DTIC FILE COPY

AAMRL-TR-89-002

AD-A206 824

DEMONSTRATION OF PHYSIOLOGICAL WORKLOAD CORRELATES IN CREW CAPABILITY SIMULATION (U)

KATHY McCLOSKEY
MELODIE MORROW
WILLIAM A. PEREZ

SYSTEMS RESEARCH LABORATORIES, INC.
A Division of Arvin/Calspan
2800 Indian Ripple Road
Dayton, Ohio 45440

JANUARY 1989

Interim Report for November 1987 to August 1988

Approved for public release; distribution is unlimited.

HARRY G. ARMSTRONG AEROSPACE MEDICAL RESEARCH LABORATORY
HUMAN SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6573

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Armstrong Aerospace Medical Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with the Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
Cameron Station
Alexandria, Virginia 22314

TECHNICAL REVIEW AND APPROVAL

AAMRL-TR-89-002

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

CHARLES BATES, JR.
Director, Human Engineering Division
Armstrong Aerospace Medical Research Laboratory

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS N/A
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AAMRL-TR-89-002
6a. NAME OF PERFORMING ORGANIZATION Systems Research Laboratories, Inc., Dayton, Ohio 45440 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AAMRL/HED
6c. ADDRESS (City, State, and ZIP Code) 2800 Indian Ripple Road Dayton, Ohio 45440	7b. ADDRESS (City, State, and ZIP Code) Wright-Patterson Air Force Base, Ohio 45433-6573
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F33615-85-C-0541
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO.
62202F	7184	10	33

11. TITLE (Include Security Classification)

Demonstration of Physiological Workload Correlates in Crew Capability Simulation (U)

12. PERSONAL AUTHOR(S)

McCloskey, Kathy, Morrow, Melodie, and Perez, William

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM Nov 87 TO Dec 88	1989 January 03	70

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
05	08	
23	01	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

Human Performance, Simulation, Workload, Heart Rate, Eyeblink, Evoked Potentials, Electrophysiology,

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A laboratory physiological measurement device was modified for use in part-mission, multicrew man-in-the-loop simulation experiments. An abstract flight simulation task was employed to obtain behavioral and physiological measures (heart rate, eyeblink, and evoked potentials) which were correlated to achieve an understanding of crew workload. The two classes of workload measures were shown to be relatable and complementary. The modified Neurophysiological Workload Test Battery device was demonstrated to be suitable for use in long duration, manned simulation experiments. Keywords

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Gilbert G. Kuperman	22b. TELEPHONE (Include Area Code) (513) 255-3727 · 22c. OFFICE SYMBOL AAMRL/HED

DD Form 1473, JUN 86 Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

SUMMARY

Behavioral and psychophysiological measures were obtained during a low-fidelity F-15 flight simulation where subjects were required to fly the wing position in relation to a canned lead flight. One of the major emphases of this preliminary research effort was to identify, solve and document the hardware/software problems that emerged when a physiological data collection device (the Neuropsychological Workload Test Battery) was interfaced with a computer controlling the simulation (Silicon Graphics device).

Another emphasis was determining the cost-effectiveness of psychophysiological measurement in terms of value of the data. This study demonstrated that heart rate and eyeblink data not only confirmed and further clarified information obtained by the behavioral measures, but also provided information about the flight task not otherwise available. It was concluded that, in future SABER simulations, the extra costs of collecting physiological data are offset by the increased dimensionality and extra information added to the flight profile data base.

PREFACE

This research was completed under contract number F33615-85-C-0541, work unit number 7184-10-33, at the Harry G. Armstrong Aerospace Medical Research Laboratory (AAMRL), located at Wright-Patterson Air Force Base, Ohio. The authors wish to thank Gilbert Kuperman, Denise Wilson and Major William Marshak, contract monitors for this project, for helpful suggestions and comments. The help of Dr Glenn Wilson on questions about the Neuropsychological Workload Test Battery (NWTB) is also gratefully acknowledged.

The authors also thank the key people from SRL who helped smooth the way. Walt Rosenthal and Barbara Richey were excellent in troubleshooting the software problems with the NWTB. Brian Porter, Curt Mayrand and Cliff Brust helped us tremendously in setting up the Silicon Graphics software. Dave Brown, Don Jones, Ken Bange and Mike Gifford were always on hand to solve hardware problems. Penny Fullenkamp and Iris Davis supported our efforts and answered questions concerning the NWTB analyses. Robyn Crawford spent many long hours with data reduction and analysis.

Without the help of these AAMRL and SRL people this research would not have been possible. To everyone, our sincere appreciation.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

AAMRL	Armstrong Aerospace Medical Research Laboratory
A/D	Analog to Digital
ANOVA	Analysis of Variance
ASCII	American Standard Code for Information Interchange
BPM	Beats Per Minute
BR	Eyeblink Rate
BV	Eyeblink Variability
DUR	Eyeblink Duration
ECG	Electrocardiography
EEG	Electroencephalography
EMG	Electromyography
EOG	Electrooculography
EP	Evoked Potentials
HDUR	Eyeblink Half Amplitude Duration
HED	Human Engineering Division
HEG	Workload and Ergonomics Branch
HR	Heart Rate
HRV	Heart Rate Variability
IBI	Inter-Beat-Intervals
IEEE	General Purpose Interface Bus (IEEE 488)
IRIS	Silicon Graphics Workstation Model
KBYTES	Kilobytes
KHZ	Kilohertz
MBYTES	Megabytes
MSEC	Milliseconds
NWTB	Neurophysiological Workload Test Battery
RMS	Root Mean Square
RMSX	Root Mean Square X Distance Between Planes
RMSY	Root Mean Square Y Distance Between Planes
RMSZ	Root Mean Square Z Distance Between Planes
SABER	Strategic Avionics Battle Management Evaluation Research
SABSRT	Software Module within the NWTB
SRL	Systems Research Laboratories, Inc.
SWAT	Subjective Workload Assessment Technique

Section 1
INTRODUCTION

The Strategic Avionics Battle-Management Evaluation and Research (SABER) facility was developed by the Human Engineering Division of the Armstrong Aerospace Medical Research Laboratory (AAMRL) to support exploratory research and development efforts concerning the evolution and refinement of advanced multiplace aircraft crew system concepts (Wilson and Kuperman, 1988). SABER exploits the responsiveness and flexibility of rapid prototyping tools and graphics display processors. SABER permits the crewstation designer/human factors practitioner to develop display formats and crew interface concepts from an abstract design to a fully interactive station, and does so through a low cost, minimum time process. Thus, SABER provides an integrated design process that begins with static display concepts and ends with demonstration/validation using full mission, man-in-the-loop simulations.

SABER was developed to provide quantitative and qualitative crew system design data early in the development process (Kuperman and Wilson, 1986). The SABER facility and process is employed to provide weapon system development agencies with data in three general areas: (1) crew workload, as a function of crew size, allocation of tasks to crew position and level of subsystem automation, (2) crew situational awareness, as a function of mission phase and external/internal events, and (3) crew training requirements, with regard to both skill requirements and proficiency levels. SABER is designed to fully capture the real-time experimental data that will support each of these three areas.

This document reports on the development and test/demonstration of one component of the SABER crew workload test capability: psychophysiological correlates of crew workload. This report presents a description of these measures, describes a version of the AAMRL Neuropsychological Workload Test Battery (NWTB) modified specifically for SABER, and presents methods and results of an off-line (non-SABER) simulation experiment which served to verify the correct functioning of the modified NWTB. It is hoped that the presentation of these data and the documentation of the NWTB modifications

will facilitate future applications of the NWTB in the SABER laboratory. (The physiological workload correlates themselves are described in the Appendix.)

Section 2
SIMULATION AND DEMONSTRATION

OVERVIEW

A small scale flight task simulation, with an NWTB (Wilson, and O'Donnell, 1988), was obtained for the SABER laboratories. The investigation had as its main objectives:

1. Test the interface between the simulation and the NWTB (identify problem areas and correct them).

2. Determine the feasibility of collecting physiological data during simulation (cost-effectiveness, value of the data, etc.).

3. Formulate recommendations for a larger simulator effort through the experience of the small scale simulation.

These three objectives were all met to some degree in the effort reported below. A description of the NWTB used to obtain physiological measures is followed by a description of the Silicon Graphics IRIS simulation (and pertinent segments of the "DOG" flight). The problems found during the interface of these two hardware and software systems (the NWTB and the flight simulation) are summarized. The experimental methods used to obtain behavioral and physiological measures during the simulation are then presented. The results of these measures (behavioral and physiological) are discussed in terms of their relation to flight parameters.

Finally, the results of the effort are used to generate recommendations for a more realistic, full scale simulator investigation that will use physiological measures as part of a larger evaluative data base.

PHYSIOLOGICAL MEASUREMENT

The NWTB and SABER Requirements

The NWTB was developed to provide multichannel physiological data collection and analysis within one self-contained measurement apparatus. In the past, collecting multiple physiological measures required multiple hardware and software configurations to handle such diverse responses as heart rate (electrocardiography--ECG), eyeblink (electrooculography--EOG), and electroencephalography (EEG). Furthermore, these separate configurations were governed by widely different signal/response characteristics. Such things as gains, offsets, sample rates, etc., varied dramatically between signals. Researchers found separate analysis of measures necessary (i.e., overlapping analysis not possible). The governing hypotheses required to make assumptions about measures and their relationship with workload also varied distinctly. Each measure required a separate "study" in both research design requirements and governing hypotheses. The NWTB was designed to eliminate a large part of these problems. The NWTB is a standardized system with well defined assumptions that can be used by laymen and physiological experts alike (Wilson and O'Donnell, 1988).

The NWTB has a PDP 11/73 central processing unit that independently controls six analog to digital (A/D) channels which are assigned as follows: three EEG channels, one ECG, one EOG, and one electromyography (EMG--muscle). The NWTB has 13 programming modules that digitize, store, and analyze data:

EEG:
1. Oddball task (audio and visual).
2. Memory scanning (Sternberg, audio and visual).
3. Continuous performance.
4. Flash evoked response.
5. Monitoring task.
6. Tracking task.
7. Brain stem response.
8. Checkerboard steady state response.
9. Sine wave steady state response.

10. Unpatterned flash steady state response.

ECG: 11. R-wave identification and time-based measurement.

EOG: 12. Blink identification and time-based measurement.

EMG: 13. Power of muscle spectra in frequency domain.

The NWTB was originally designed for collection of data from only one operator for limited periods of time (1 hour). This 1-operator/1-hour limitation was not acceptable for use in the SABER laboratories. SABER requirements were such that data collection included a 2-operator scenario for a period of up to 6 hours. Furthermore, physiological data were to be digitized and stored during the entire 6-hour period (the original NWTB stored data in discrete time blocks). This not only required huge amounts of memory to be added to the NWTB capability, but also extra programming modules that allowed marking of the time series and transfer of the pertinent data segments from memory to the original NWTB format for analysis.

The following section of this report discusses in detail the problems of interfacing/integrating the NWTB with an off-line simulation that possessed requirements identical to those of the SABER laboratories.

The Modified Neuropsychological Workload Test Battery

The NWTB was modified to accommodate studies in the SABER laboratories, in which lengthy data collection could last up to 6 hours for two subjects at a time. A two-step approach was employed. First, the problem of data collection was addressed.

Because a large quantity of data could potentially be collected, 8 more disc surfaces were added to the NWTB to increase storage capability to 120 Mbytes. Ten analog to digital (A/D) channels were available for collecting and storing data. Two of the 10 channels were dedicated to heart rate collection and sampled at a 1 KHz rate (4 Kbytes/second, 240 Kbytes/minute, 14.4 Mbytes/hour, and 86.4 Mbytes/total test capability). Two channels were dedicated to eyeblink collection and sampled at a 100 Hz rate (8.64 Mbytes/total test capability). Six channels were dedicated to EEG collection and sampled at a 200 Hz rate for intermittent 3-minute periods

(432 Kbytes/period and 17.28 Mbytes/40 periods). Total data collection capability for all 10 channels was 112.32 Mbytes. (NOTE: The actual memory space required for the present study was negligible and did not approach the maximum capacity of the modified NWTB; for each subject, one channel of heart rate, eyeblink, and EEG data were collected for very short periods of time, i.e., under 1 hour continuous for one subject.)

The next step was to arrange the data into a format that could be presented to the standard NWTB hardware and software for analyses. Extra A/D converters were added to the NWTB to accommodate the two channels of heart rate data. A software module (also called SABER) was written to collect heart rate and eyeblink data. This module also provided an "Audio Rare Event" (Oddball) task, described later, identical to the standard test on the original NWTB. The module transferred all collected digitized data to the hard discs for archival storage. This feature allowed previously collected test data to be restored in the modified NWTB, ready for the next processing step. An IEEE interface between the NWTB and SABER simulation was provided for a direct link concerning control and data transfer communications but was not used in the present study. A second software module (SABSRT) was created to allow the operator to select position and time frames of physiological data that corresponded to specific flight segments. These "data windows" were identical to standard NWTB "*.DAT" files, which allowed the information to be processed in the NWTB data reduction/analysis program as usual. A third software module controlled the gain and filter settings of a Systems Research Laboratories, Inc. (SRL), programmable amplifier/filter. The amplifier/filter settings used in this study were as follows:

	EEG	EOG	ECG
High Pass:	0.10	0.10	9.65
Low Pass:	29.80	100.44	100.44
Filter Range:	32-64K	4-8K	1-2K
Gain:	50,000	5000	2000
Notch Filter:	60Hz	60Hz	60Hz

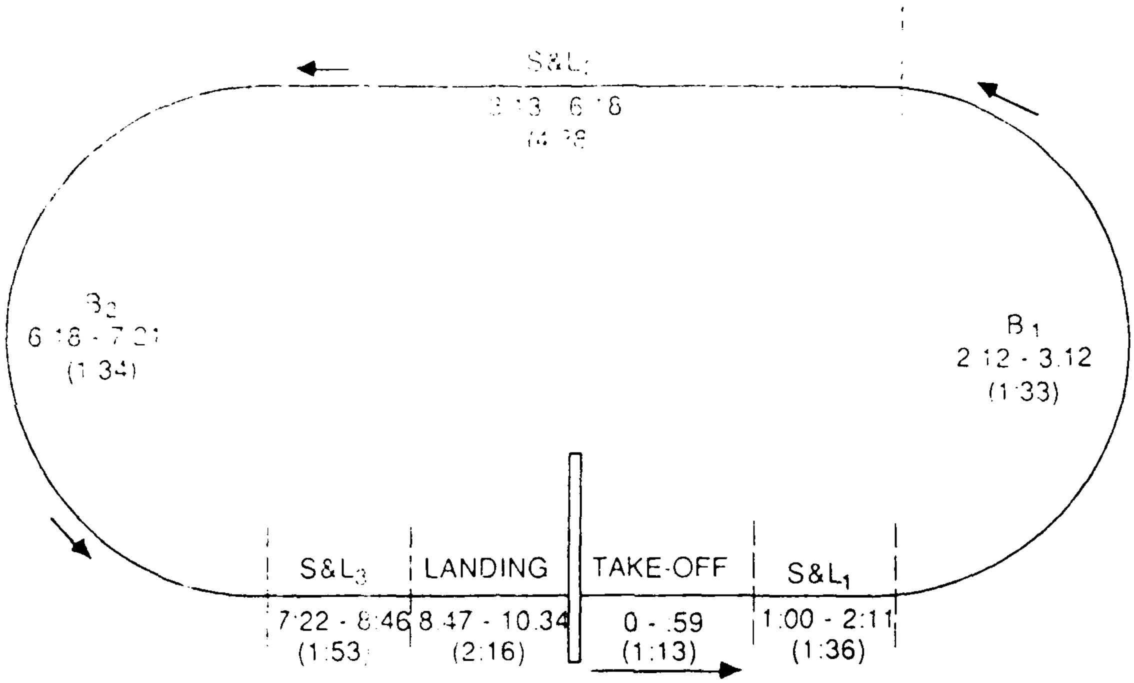
In summary, hardware additions included 8 disc surfaces, A/D converters and a 10-channel SRL amplifier/filter. The sequence of events was as follows: data were collected on the NWTB as usual, however, the digitized data went directly to the 8 disc surfaces that had been added to the NWTB. After data collection was completed, SABSRT pulled windows of data from the discs. Once chosen, these windows went directly onto the standard NWTB removable 5 Mbyte disc as "*.DAT" files where they could be examined with the standard NWTB reduction/analysis program modules.

SIMULATION

IRIS Silicon Graphics Computer System

A dynamic flight mission using a modified version of the "DOG" program provided the task. The DOG software included an option for recording the performance parameters (e.g., altitude, velocity) at selectable data sampling rates. The hardware of the Silicon Graphics Workstation Model (IRIS 3030) remained standard. A flight pattern similar to a simple oval circuit was created with a simulated F-15 aircraft and saved (prerecorded) on the IRIS' hard disc. This prerecorded flight path was sequenced in the following manner: take-off (TO), straight and level (SL1), left bank (B1), straight and level (SL2), left bank (B2), straight and level (SL3), and landing (L). An illustration of the flight path is presented in Figure 1. Each segment of the pattern used at least one minute's worth of time to enable EEG data collection. Next, the pattern was broken down into actual time windows that reflected each flight segment. Once the criterion flight pattern was prerecorded and timed, subjects piloted a second F-15 aircraft along with the prerecorded criterion aircraft. The objective was to fly the second (wing) aircraft 1500 feet or less behind the first (lead) aircraft, maintaining similar altitude and roll angle. Unfortunately, the IRIS update rate slowed down when the wing plane was added to the screen with the prerecorded lead plane, and sampling rate decreased. Sampling rate also became more variable within each segment, and within and between each subject. When a sampling rate was specified, the display made jerky, discrete movements instead of the smooth, continuous movements required for simulation. Because this "jerky" simulation was too difficult to fly,

TOP VIEW OF MISSION WITH SOFTWARE TIME (REAL TIME)



S&L - STRAIGHT AND LEVEL
B - BANK

SIDE VIEW OF MISSION

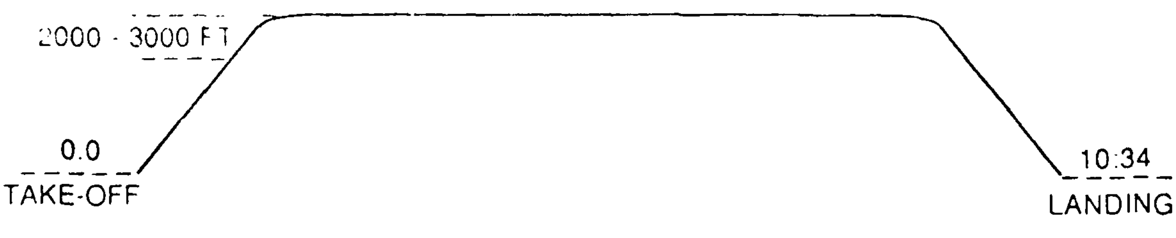


Figure 1. Flight Path Top and Side Views

sampling rate was allowed to vary randomly, resulting in about 13 performance samples per second.

Specifying the lead and wing options of the IRIS allowed the subjects' wing flight and the prerecorded lead flight to be recorded in a file simultaneously, resulting in comparable data at the end of a run. The data showed both aircrafts' positions in the Z (longitudinal trailing distance), Y (altitude), and X (lateral off-set or roll) axes for each sample taken (Figure 2). Thus, flying "two" was similar to a three axis pursuit tracking task.

Two programs, FLIGHT.C and COMM.C, which encompass the DOG program, underwent modifications to allow a distance and heading display to be placed on the screen. Also, an additional modification of COMM.C allowed position data to be sent to "OUTFILE" for viewing.

NWTB AND IRIS INTERFACE

The new problem of the random IRIS sampling rate during simulated lead/wing flight had implications for the use of the NWTB in the study. The IRIS ran on variable software time instead of real-time. The NWTB ran on a real-time clock, and problems appeared during correlation of NWTB and IRIS data. To solve these problems, an RS-232 was employed to link both systems together to run on the IRIS' software time. Since the software time was considerably slower, the segments lasted longer than we had originally anticipated. The segment start and stop times, as well as the amount of real time, are listed below.

	Prerecorded Time Start	Stop	Real Time
Take-off	00:00	00:59	01:13
SL1	01:00	02:11	01:36
B1	02:12	03:12	01.33
SL2	03:13	06:18	04:38
B2	06:19	07:21	01:34
SL3	07:22	08:46	01:53
Landing	08:47	10:34	02:16

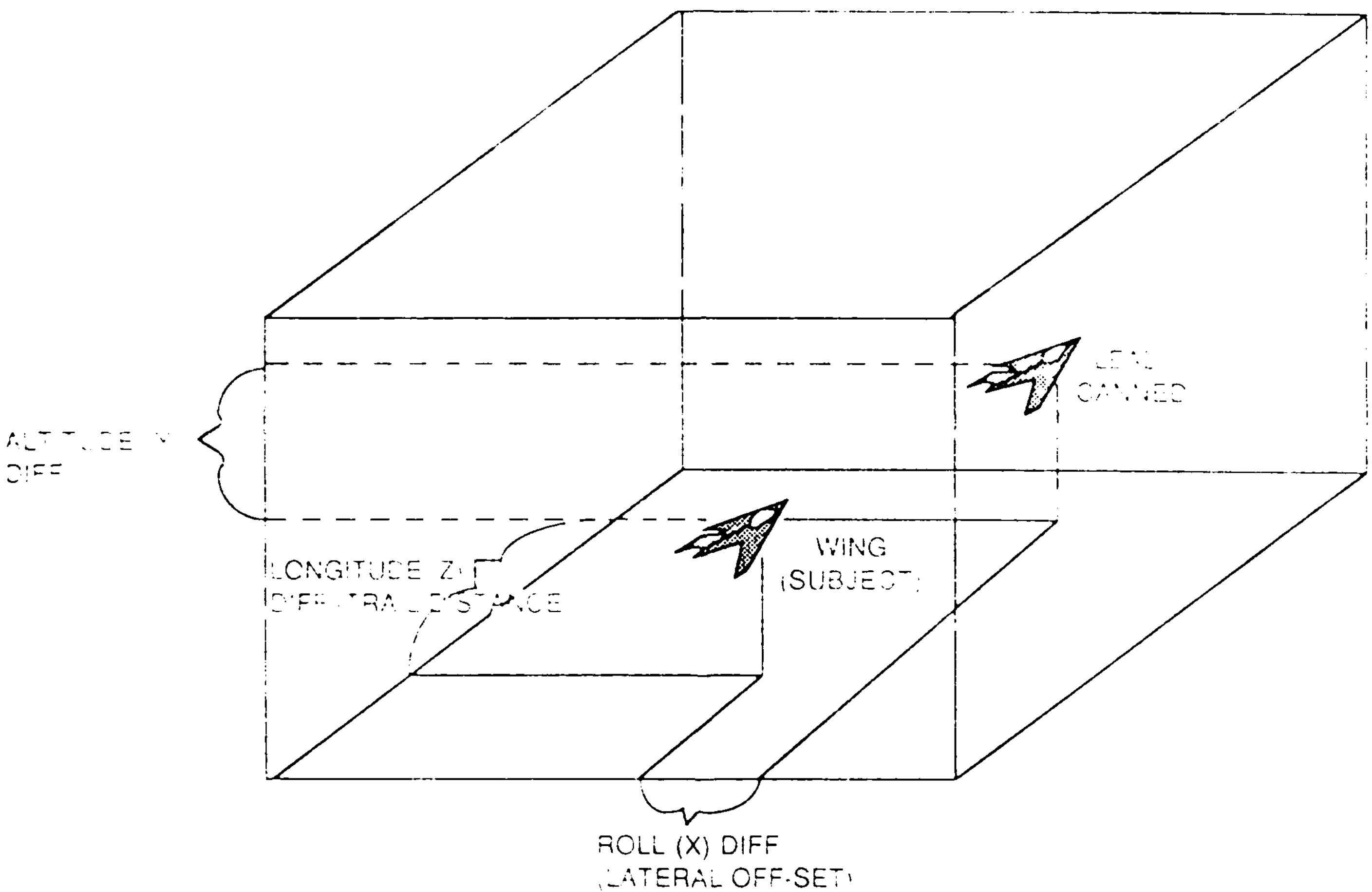


Figure 2. Representation of the Lead and Subject Flight in the Three Dimensions

Although the real start and stop time for each of the flight segments was known, the prerecorded time was variable in both interval between seconds and length of each second. Because the NWTB used the IRIS' variable software time, the NWTB could be programmed to begin data collection at 00:00:00 when the IRIS clock began counting at the start of the flight trial. For the EOG and ECG, this was not a problem since the data was recorded continuously and setting the program parameters to collect this type of data was a one-step operation. However, for the EEG data a real logistics problem emerged. During EEG collection, a secondary task (described later) was presented and evoked potentials recorded to the onset of the task items. A window of time during each segment was chosen to present this secondary task, rather than attempting to collect continuous data. This meant that each time window block had to be typed in separately. Unfortunately, the operator could only input one window at a time while collection was underway. This was an extremely inconvenient and error prone situation. If the time window block was entered incorrectly, and could not be corrected by the time the test was due to begin, the entire run had to be reset and restarted. To alleviate this problem, the NWTB was programmed to start counting from 10:00:00 to allow the operator to type in all the time window blocks before the run started. When the task started, the NWTB time was reset to 10:00:00 and the EEG data was collected at time points specified by the operator. (NOTE: All the data windows were under 00:10:34, so it would have taken the NWTB almost 10 hours to count down to the first time window.) The starting times for each data window were input as follows:

	EOG and ECG	EEG
Take-off	10:00:03	10:00:03
SL1	10:01:00	10:01:07
B1	10:02:11	10:02:11
SL2	10:05:15	10:05:15
B2	10:06:19	10:06:19
SL3	10:07:22	10:07:22
Landing	10:08:47	10:09:25

EEG windows were slightly different from EOG and ECG because the data in each window were collected as discrete data. The window choices were made to optimize the effects of each segment. Each 1 minute EEG sample needed to reflect the effects of each segment. For example, "landing" effects would most likely be seen in the latter part of the landing segment where the subject's plane was actually touching down on the runway. Once the run was completed (when the lead plane stopped on the runway) the NWTB program was halted and SABSRT was used to sift through the data for the specified windows listed above.

Section 3
DATA REDUCTION

REDUCING THE IRIS BEHAVIORAL DATA

After the behavioral data was sampled in binary fashion, a program was written to convert the data set into American Standard Code for Information Interchange (ASCII) format. This same program further reduced the data into root mean square (RMS) error terms for each of the three axes (X, Y and Z). The RMS formula used here is presented below:

$$\text{RMS Error} = \sqrt{\frac{E(X_{Li} - X_{Wi}) - \dfrac{(E(X_{Li} - W_{Wi}))^2}{n}}{n - 1}}$$

L = Lead (canned)
W = Wing (subject)

The data were sorted between subjects and by flight segment. The data set was then written onto an ASCII IRIS.* file which was compatible with SAS. These files were transferred to a VAX 8650 via a Telnet communications interface. Once on the VAX the data set was ready to be statistically examined.

REDUCING THE NWTB PHYSIOLOGICAL DATA

Analysis of the physiological data on the NWTB was a very time consuming task during this study. The ECG, EOG, and evoked potentials required separate analysis techniques. Each of these techniques is described below.

ECG Analysis

The heart rate routine calculated the time between R waves, or inter-beat-intervals (IBIs), and the variance of the IBIs. Prior to these calculations, the program allowed certain parameters to be manipulated by the

operator. These parameters included the minimum and maximum IBI time values, cardiac (R wave) amplitude, and a difference criterion in terms of slope of the R wave. For most of the subjects the default parameter values provided by the program were sufficient for the correct identification of the R waves, and the resultant IBIs. By calling a file name within the heart rate program, the first 10-second block of heart rate data was displayed on the terminal. The R waves on the screen were marked with an asterisk by the program. The operator visually inspected the R waves and when it was determined that the program was identifying them correctly, a "C" was entered on the keyboard. Then the program calculated statistics for the entire file. The program displayed these statistics on the screen and noted any "bad beats" (those falling outside of the parameters) for the file. In most cases, adjusting the parameters removed the bad beats. Yet a problem arose when R waves occurred on the "edges" (beginning or end) of the 10-second blocks. These R waves were lost to the analysis and introduced artifactual variance to the IBI averages. This same problem was outlined by Albery (1988). (It is suggested that an option be added to the heart rate program that allows the operator to insert an asterisk at the correct R wave occurrence, even if it falls at the edge of a 10-second block.)

The NWTB allowed for a printout of the summary statistics via a printer mounted at the front of the computer. This capability ensured that the operator obtained hard data copies of each file summary. However, the NWTB did not possess the capability to organize multiple heart rate files into a "master" file for further analysis. The summary statistics were entered by hand onto a data sheet, and ultimately input into a SAS file. (It was recommended that this labor intensive process be replaced by a software module that will organize master files.) Recently, developing an NWTB interface for file transfer to other computers that have statistics software has greatly reduced operator time. Unfortunately, this interface was not available for use at the time of our data transfer.

The NWTB summary statistics gave the averages of the IBIs for each of the 10-second blocks, as well as the overall segment average. When it was decided post hoc to break the overall segment averages down into the first

and last 30 seconds of the segments, the IBI values were averaged across the first three and last three 10-second blocks. The variability of the IBIs could not be averaged by blocks nor be examined in the same manner as the IBIs. (A small program module that would allow different combination averages of the 10-second blocks is another recommended option that would have reduced operator time and produced a more detailed analysis.)

EOG Analysis

The eyeblink data were the easiest to reduce using the NWTB. The feature that expedited the analysis was that of a moving cursor under the control of the operator. The parameters that identified an eyeblink were those of blink amplitude, minimum and maximum blink time and slope of the eye closure. Usually these parameters were sufficient. At times, however, blinks would occur that did not fit the subject's overall blink pattern. Changing the parameters to include or exclude an unusual blink often affected more "normal" blinks. The cursor option eliminated this problem by allowing the operator to add or delete a blink without changing parameter values. As with the heart rate program, when the operator was satisfied that the program was identifying blinks correctly, a "C" was entered and summary statistics obtained. A hard copy was also available from the NWTB printer. The form of the summary statistics was different in that number of blinks, inter blink interval, closing duration, and one-half amplitude closing duration was given for the eyeblink data.

The same problems occurred during the eyeblink analysis as those found with the heart rate program, specifically, entering data by hand and lack of post hoc analysis programs. When examining the first and last 30 seconds of each segment, only the number of blinks could be averaged across the 10-second blocks.

Other problems emerged during the eyeblink analysis not found with heart rate. In some instances, subjects did not blink at all. The operator was confronted with zeros across all summary statistics, and was required to scroll through the entire data file visually to ensure that the program had not missed real blink occurrences. Furthermore, two or more blinks had to

occur within a 10-second block, or the blink variance could not be calculated. This was not a problem unique to the NWTB. Missing data (no blinks, or less than two blinks) are a function of the eyeblink measure, and were treated here as pertinent data during later analysis.

Evoked Potential Analysis

The routine on the NWTB averaged together the single evoked potentials elicited by the tones. The program produced separate averages for rare and frequent tones. Furthermore, the program automatically placed a vertical line at the place on the evoked potential (EP) averaged waveform that was the largest in amplitude. The NWTB also provided an operator controlled cursor that allowed amplitude and latency measures to be obtained for any of the components of the averaged EPs.

The two most time consuming tasks in removing the pertinent data from the NWTB were those of "picking" the correct components of each waveform by moving the cursor, and transferring the corresponding amplitude and latency values to data sheets by hand (and ultimately to another computer for statistical analysis).

Other problems were encountered that were unique to the evoked potential measures. One minute of EP recording during the flight task limited the number of single trials available for each average. This problem was further compounded by the subjects' tendency to blink at the onset of each tone, especially the rare tones. Without an eyeblink correction program on the NWTB, the operator had to determine the minimum eyeblink amplitude threshold for each EP average, without going below a 12 single trial cutoff for each average. The overall eyeblink amplitude thresholds for the EP averages ranged from 500 microvolts to two and one-half millivolts.

Section 4
EXPERIMENTAL METHODS

SUBJECTS

Six healthy male nonpilots, aged 20-45 years, participated in this study. Each subject was right-handed and had approximately 20/20 uncorrected vision.

DESIGN

Data were collected in a mixed factorial design. Independent measures included training session (early, middle or late training, test session 1 or 2), visibility (day or night), and flight segment (REST, TO, SL1, B1, SL2, B2, SL3, and L). The visibility factor was counterbalanced within and between subjects. Night was the same flight scenario as day, but the display became dimmer and "night lights" from the city, runway, and aircraft became apparent. Dependent measures included root mean square error, which represented distance between the lead/wing planes in the three axes (XRMS, YRMS, and ZRMS), heart rate and heart rate variability (HR and HRV), eyeblink rate, duration, half amplitude duration and variability (BR, DUR, HDUR and BV), and N200/P300 amplitude and latency of the evoked potentials obtained from the EEG.

PROCEDURES

Training

The primary task required the subjects to follow the prerecorded lead flightpath, described previously. Training was composed of 12 sessions which lasted for 1 hour and 15 minutes each. On the first day, subjects were introduced to the task and given the opportunity to operate the simulated F-15 aircraft. On the second day, they learned how to fly behind the prerecorded lead aircraft (see Table 1 for the training schedule). Following these 2 days, subjects received standardized training which consisted of six full missions and four "landing only" segments. Following

TABLE 1. TRAINING SCHEDULE SHEET

Subject ________ Condition Order ________ Date ________

DAY 1:
Familiarization:
- Experimental Purpose and Task
- Introduction to the NWTB
  - Purpose
  - Instrumentation
  - Data Collection
- Introduction to Flight Program
  - Flight Path
  - Flight Procedures
  - Briefing Demo
- Sample Physiological Data
- Focused Training
  - Overview of Controls
  - Flaps and Spoilers
  - Demo Flight
- Hands On Training
  - Day and Night
  - (D) (N)
    - Take-Off Procedure for Lift Off
    - Straight and Level Altitude and Jink
    - Banking Procedure for Turning
    - Landing
      - Approach
      - Speed
      - Rate of Descent
      - Controls
      - Touchdown

DAY 2:
Review:
- Flight Procedures
- Overview of Controls
- Flaps and Spoilers
- Day and Night Flights
- Introduction to Top Gun
  - Formation Flying
  - Demo of Prerecorded Flight
  - Performance Requirements
  - Following Lead
- Hands On Wing Training (same as DAY 1 Hands On Training)

the second (early), seventh (middle) and twelfth (late) training sessions, all the dependent measures were collected except the EEG, to provide information on training progress. On these data collection days, a 2-minute EOG and ECG baseline measurement was taken, followed by two full warm-up flights. Following warm-up, two full missions were flown where EOG, ECG and RMS error scores were collected.

Training criterion was reached when subjects could maintain a distance of 1500 feet maximum between his own plane and that of the lead. A landing with a score of zero (no crash) had to be obtained at least half the time. Of six subjects, five reached criterion by the last training session. These five subjects' data were used in all subsequent analyses.

Testing

Data were also collected on the sixteenth (test 1) and the seventeenth (test 2) day. Because EEG data were collected during these sessions, which resulted in longer preparation and data reduction times, each session lasted approximately 3 hours. Similar to the data collection for training, two minutes of EOG and ECG baseline data were collected prior to flight. Also, EEG data were collected during the second minute prior to flight while the subjects monitored audio rare events (see below). Following baseline, subjects were given two full warm-up flights, and then required to fly two full missions during which all dependent measures were collected. The audio rare event occurred for 1 minute during each segment for both test 1 and 2 lights.

Audio Rare Event (Oddball)

The secondary task was the Audio Rare Event, or Oddball test, which was used to elicit evoked potentials from the background EEG. This standard test was taken from the NWTB. The Oddball intermittently presented two tones of different pitch to the subject via headphones. The subject's task was to count the number of tones occurring in a specific pitch while ignoring the other tone. The task was designed to present the counted tone

between 20 to 40 percent of the time. Hence the name "rare" and "oddball." The parameters for this test are presented in Table 2.

TABLE 2. PARAMETERS FOR THE AUDITORY ODDBALL

Rate tone:	1501 Hz
Frequent tone:	1199 Hz
Interstimulus interval:	1500 msec
Probability of rare tone:	40 percent
Minimum number of rare tones:	14
Tone intensity:	1 (approx. 75-80 Db (A))
Test will last 60 seconds *	

* (machine time was actually longer)

Electrode Placement and Procedure

Silver-silver chloride electrodes were used for data collection. Adhesive collars were placed around the cuff surrounding the electrode. The electrode was then filled with conducting cream. The subjects' skin was lightly scrubbed with a mild abrasive gel, rinsed with alcohol, and dried with a gauze pad. Following preparations, one electrode was placed on each of the subjects' mastoids, one for reference and one for ground. For EOG, an electrode was placed centrally just above the eyebrow of the dominant eye. For ECG, an electrode was placed one-half inch above the top of the sternum (fleshy midline indentation), with a reference electrode placed on the subject's side, just below the lowest left rib. For the EEG, an electrode was positioned on the central parietal area (Pz) according to the 10-20 system of scalp electrode placement (Jasper, 1958). Impedances between signal and reference electrodes were kept below 20K ohms for ECG, below 10K ohms for EOG and below 5K ohms for EEG. The mastoid electrode with the lowest impedance was used to reference the EOG and EEG; the other electrode was used as ground for all three signals. Impedance for all connections was checked periodically throughout the data sessions.

Section 5
RESULTS

The following paired comparisons for the behavioral data were performed using standard one-tailed t-tests, significant at $p < 0.05$. The paired comparisons for the physiological data were performed using the more conservative Bonferroni test, also significant at $p < 0.05$. The analysis of variances (ANOVAs) reported here were obtained through the 1985 version of SAS.

BEHAVIORAL DATA

The X axis data reported below represents the amount of the subjects' lateral offset compared to the wing position. The Y axis data corresponds to the altitude difference between the subject and lead. Finally, the Z axis corresponds to the trail distance of the subject to the lead.

X Axis (Lateral Offset)

RMSX. The root mean square error in the X axis (RMSX) was affected by flight segment, $F(6,24) = 15.32$, $p < 0.0001$. As can be seen in Figure 3, RMSX was larger during the two bank segments than during the rest of the flight. Day/night conditions and session block mediated this effect. The segment by day/night interaction was significant, $F(6,24) = 3.34$, $p < 0.0154$. This interaction is depicted in Figure 4. During the first bank segment, RMSX was larger during the day condition than during night. The segment by session block interaction was also significant, $F(24,9) = 3.44$, $p < 0.0001$. As can be seen in Figure 5, during the first and second bank segments and second straight and level the RMSX was larger early in the session blocks.

Y Axis (Altitude).

RMSY. The only significant effect for RMSY was that of session block, $F(4,16) = 3.54$, $p < 0.0298$. This effect is shown in Figure 6. RMSY was larger during the early session block than during the rest of the sessions.

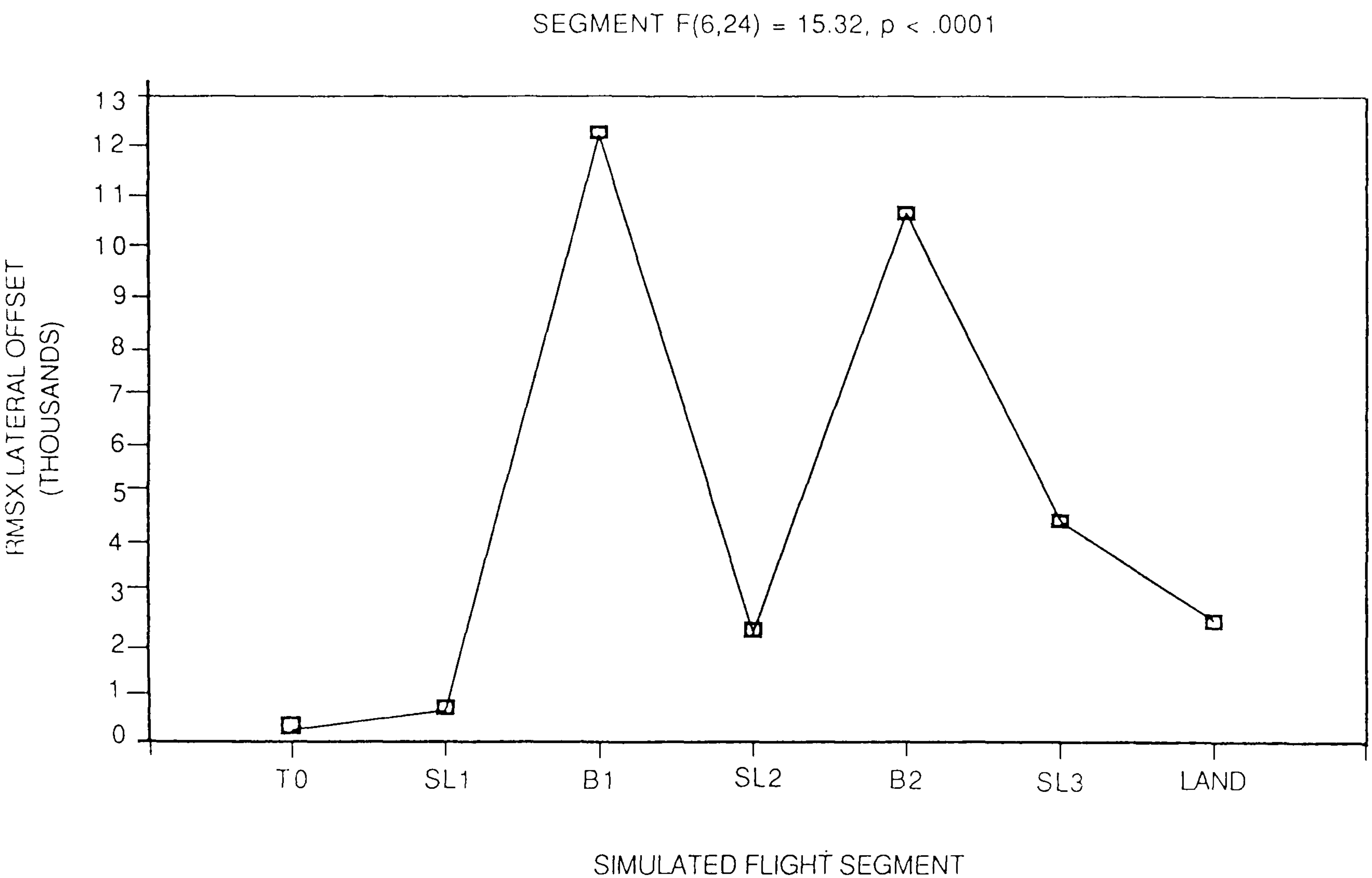


Figure 3. RMSX Lateral Offset Segment Effects

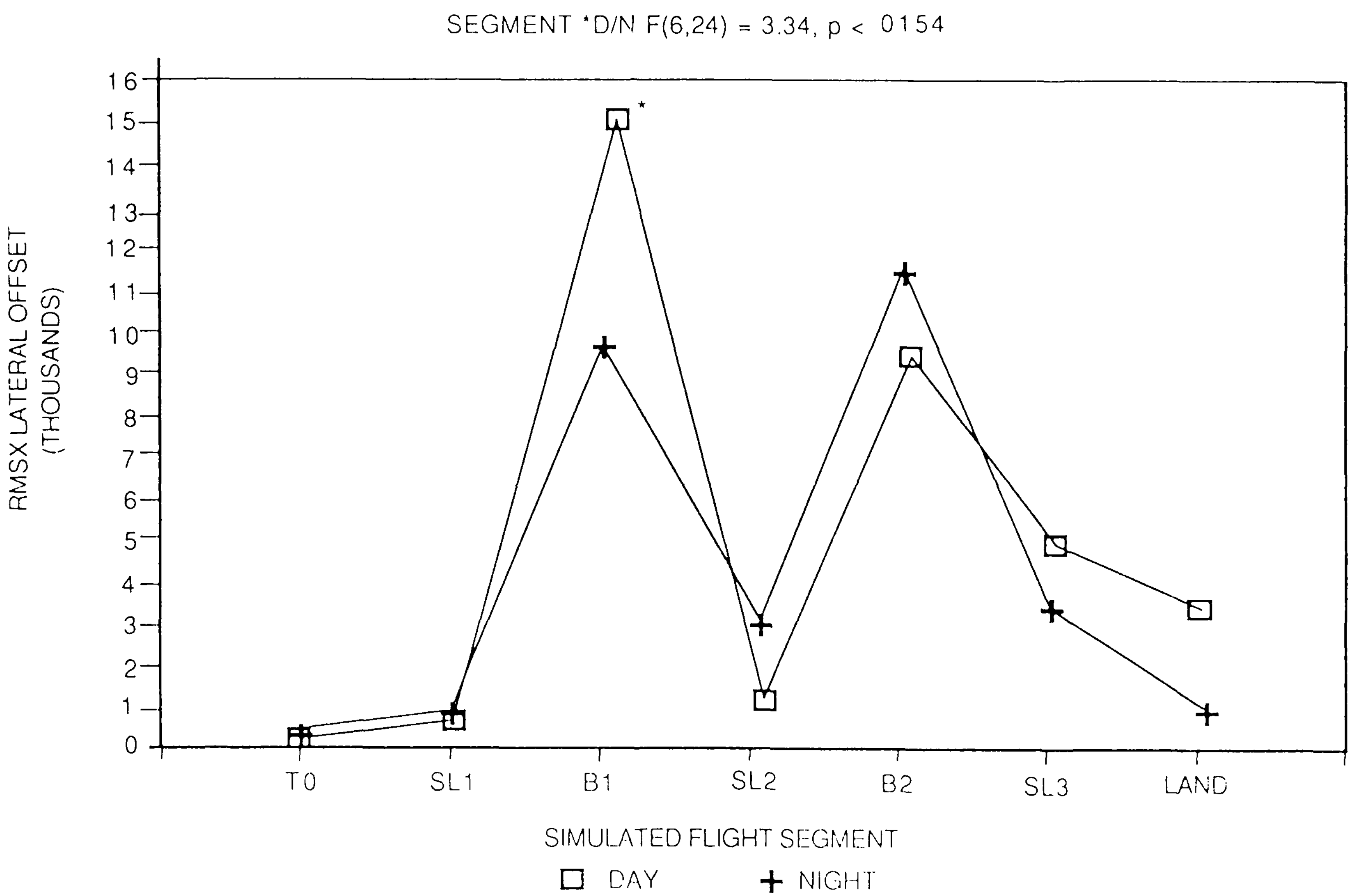


Figure 4. RMSX Lateral Offset Segment by Visibility Effects

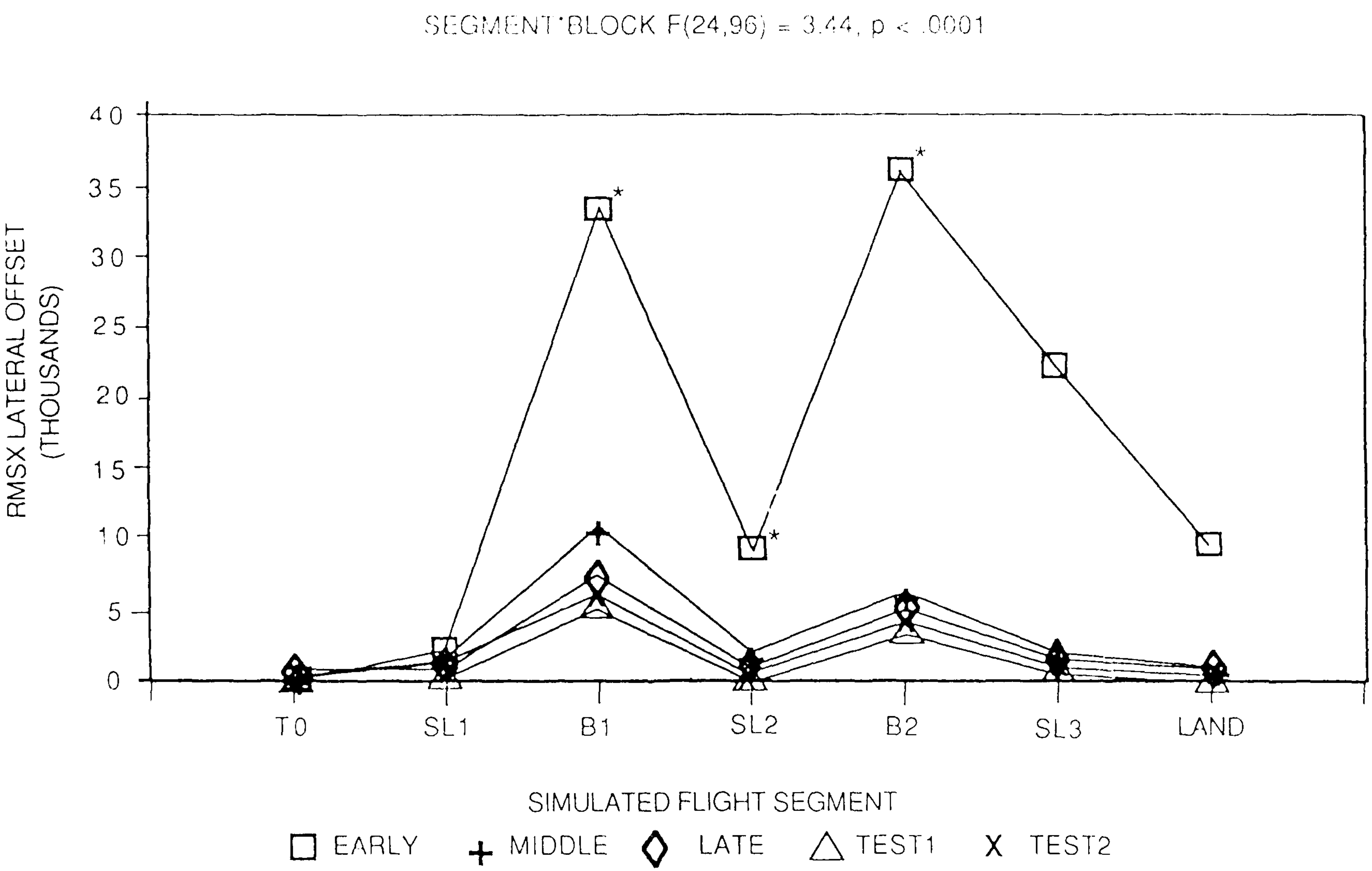


Figure 5. RMSX Lateral Offset Segment by Session Block Effects

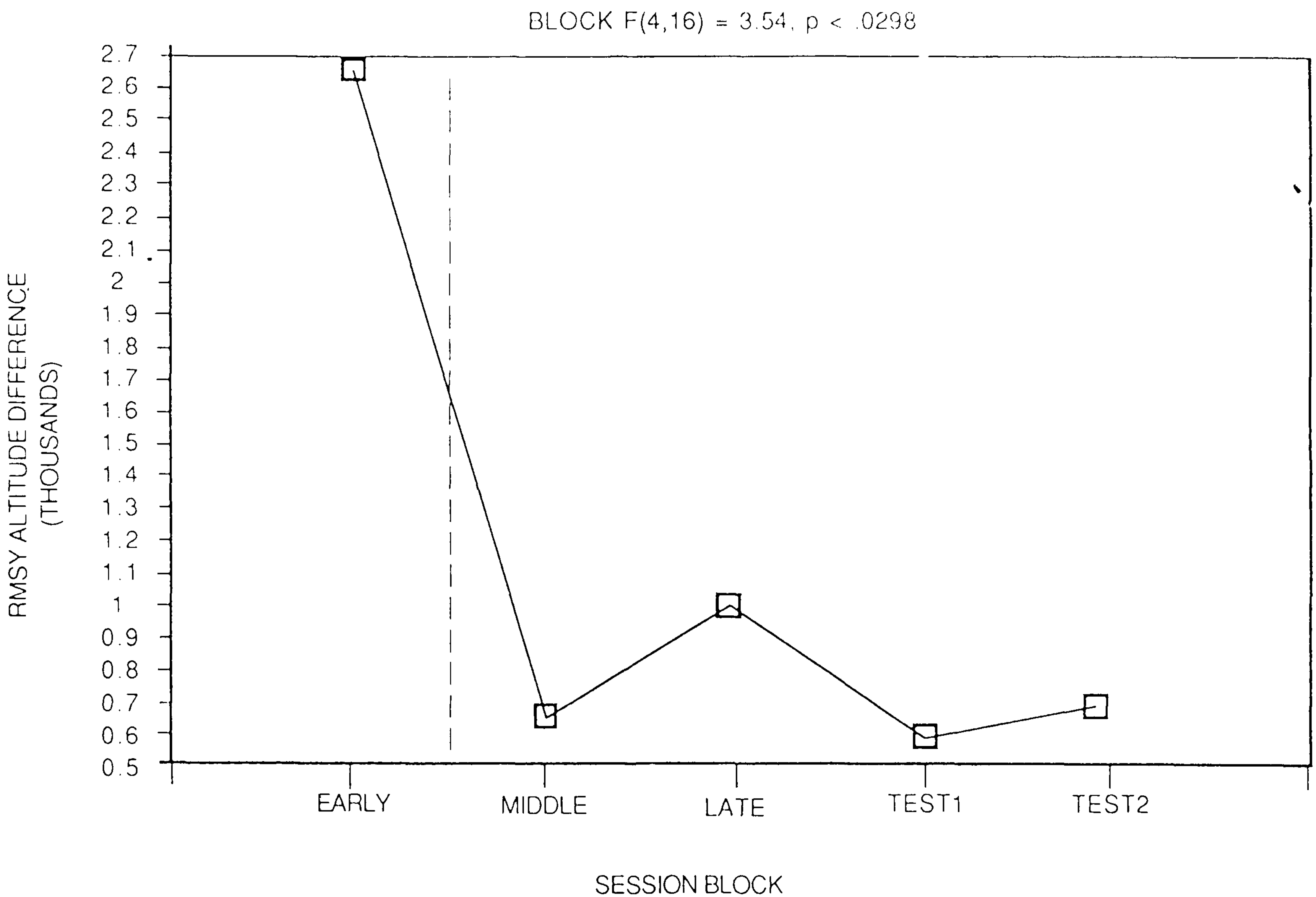


Figure 6. RMSY Altitude Session Block Effects

Z Axis (Trailing Distance)

RMSZ. There was a significant effect of segment on RMSZ, $F(6,24) = 3.11$, $p < 0.0213$. RMSZ was larger during the second and third straight and level and the second bank than during the other flight segments (Figure 7). Similar to the X axis, this segment effect was mediated by day/night conditions and session block. The segment by day/night interaction was significant, $F(6,24) = 2.65$, $p < 0.0410$. As depicted in Figure 8, RMSZ was larger at the second straight and level segment during day than during night. The segment by session block interaction was significant, $F(24,96) = 2.27$, $p < 0.0026$. Figure 9 shows that RMSZ was larger at the second and third straight and level, the second bank, and landing segments during day flight than during night.

PHYSIOLOGICAL DATA

Heart Rate (ECG)

Beats Per Minute (BPM). The first test performed on the heart rate in BPM was a flight segment by session block interaction ANOVA (8 x 5). This interaction was not significant. The main effect of segment was significant, $F(7,28) = 3.97$, $p < 0.0039$. As shown in Figure 10, during the landing segment subject's heart rate was larger than during rest, take-off, the first straight and level and bank, and the second straight and level. The second bank and third straight and level were not significantly different from landing. The main effects of session block and day/night were not significant.

Another test of heart rate was performed on BPM taken from the first and last 30 seconds of each segment. This flight segment by time block analysis of variance (ANOVA) interaction was not significant, even though visual inspection of Figure 11 suggests that BPM is larger during the last 30 seconds of the landing than during the first 30 seconds.

Even though the main effect of session block was not significant, a pattern of the five subjects' means emerged along training order. A regression was

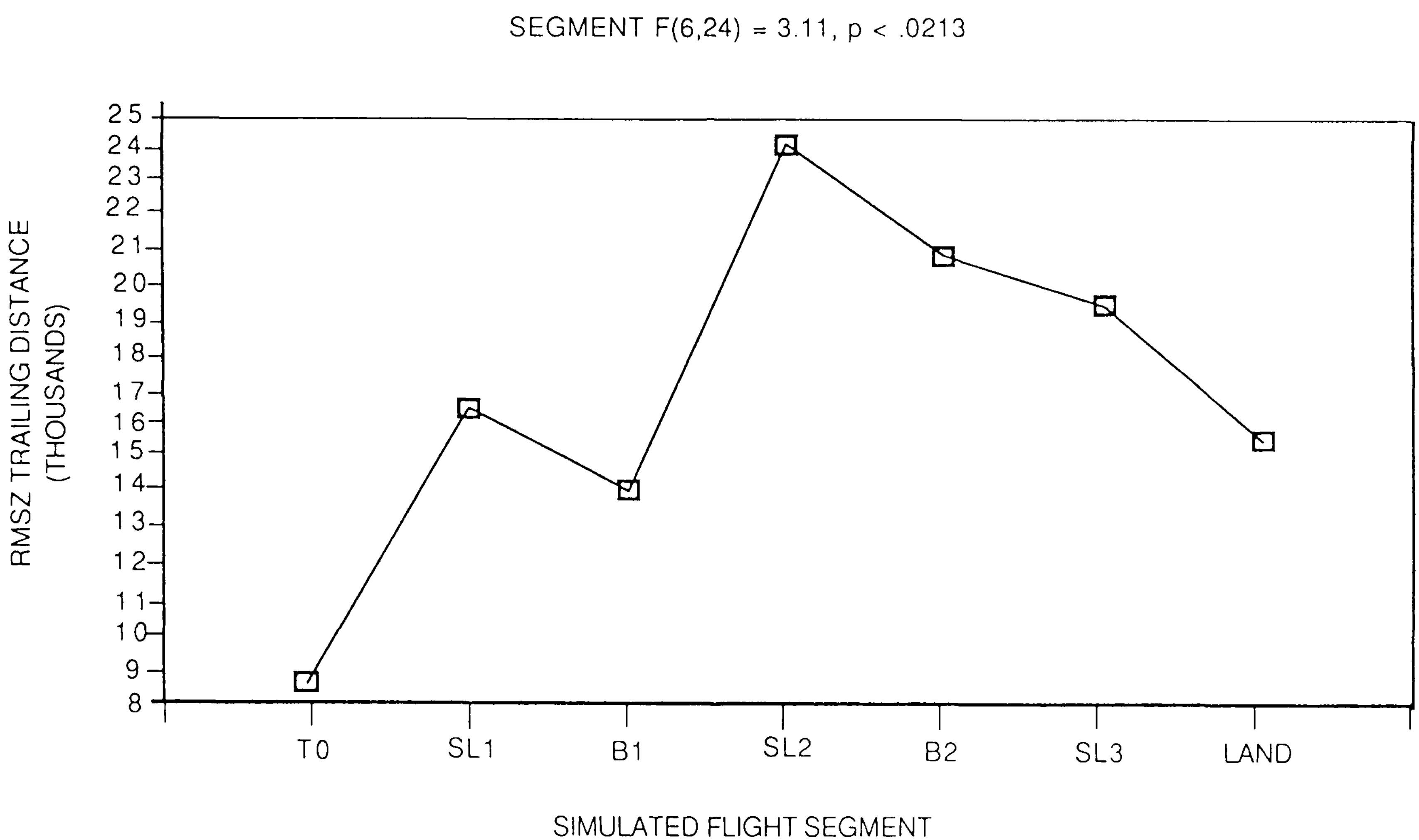


Figure 7. RMSZ Trail Distance Segment Effects

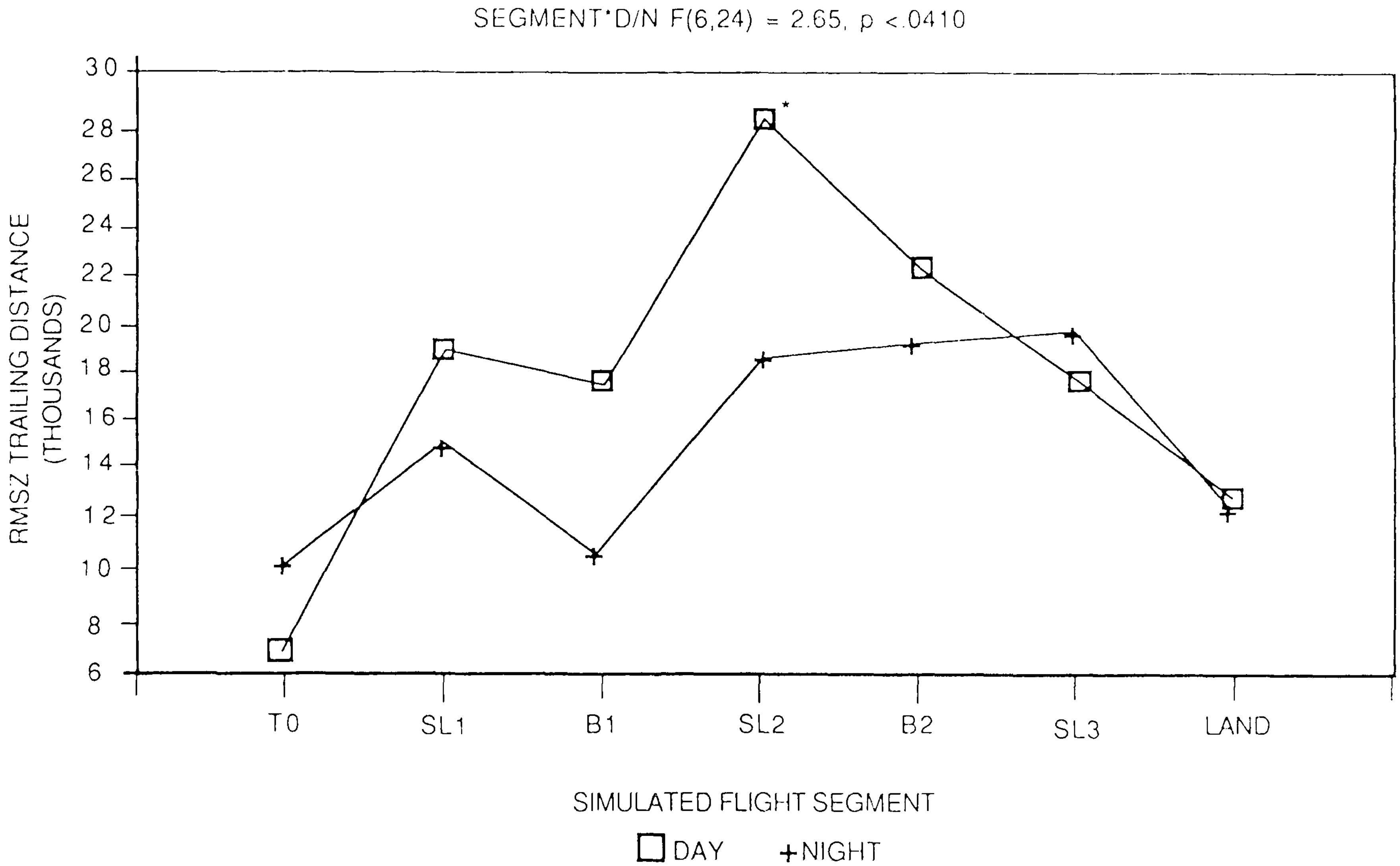


Figure 8. RMSZ Trail Distance Segment by Visibility Effects

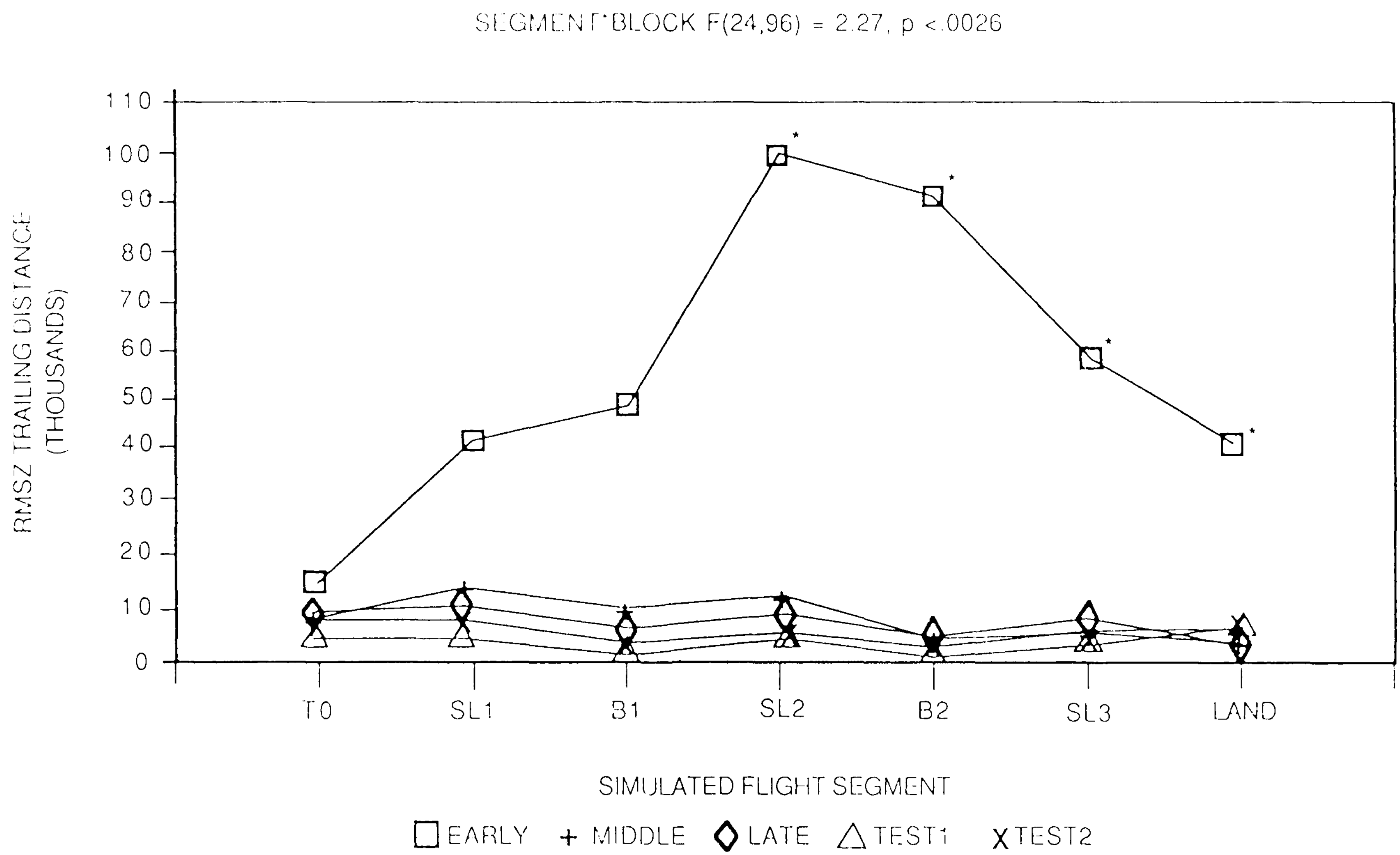


Figure 9. RMSZ Trail Distance Segment by Session Block Effects

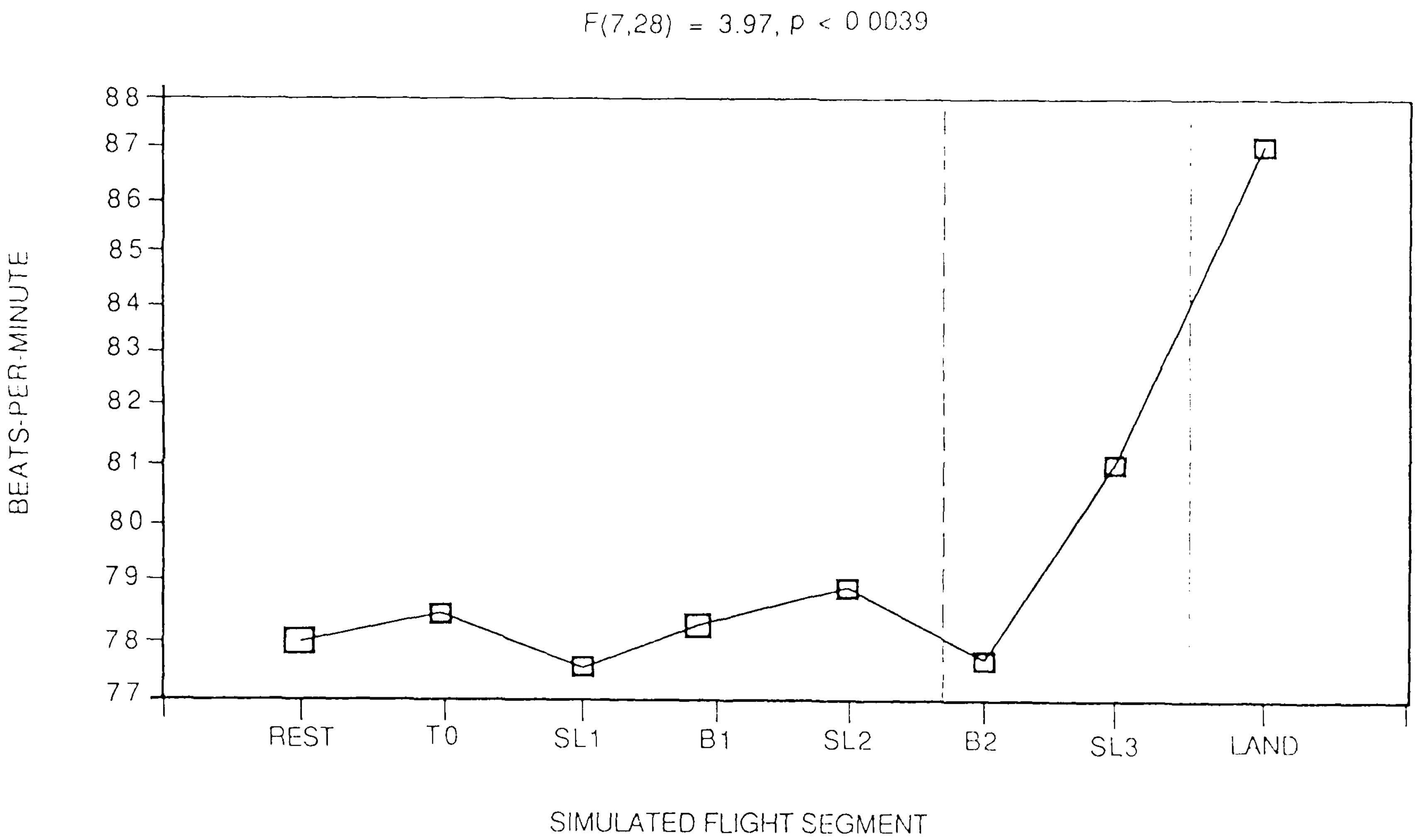


Figure 10. Beats Per Minute Segment Effects

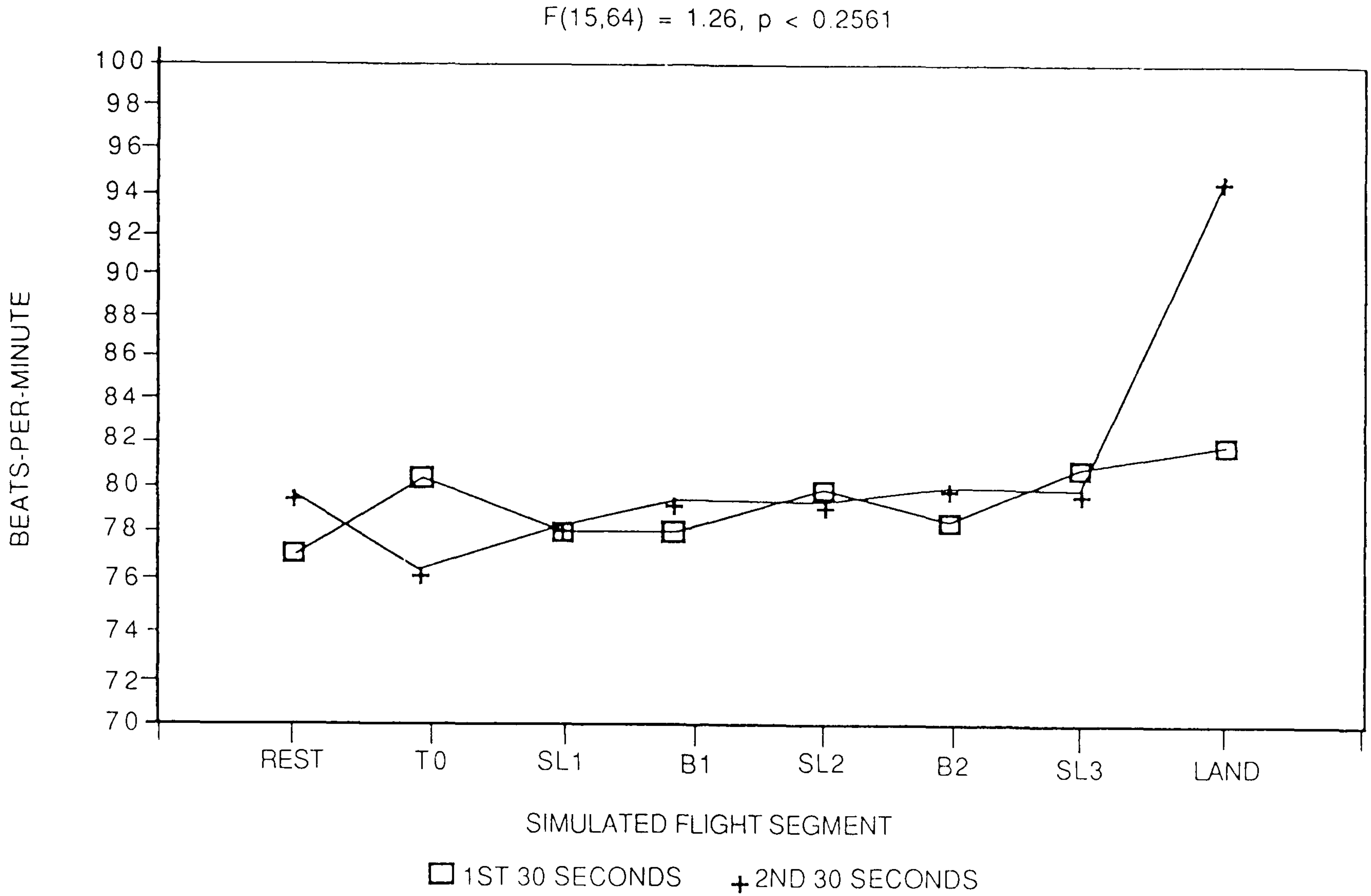


Figure 11. Beats Per Minute Segment Effects in First and Second 30-Second Blocks

performed on the means, as shown in Figure 12. The r value equaled 0.7743, yet accounted for only 59.9 percent of the variance. Again, a pattern is apparent upon inspection, but not statistically significant.

Heart Rate Variability (HRV). The first test performed on the HRV was also a flight segment by session block ANOVA, which was not significant. All main effects, day/night, flight segment and session block, were not significant. However, HRV in relation to flight segment is presented in Figure 13. Recall the behavioral data in the X axis (lateral offset) during the bank segments (Figures 4, 5 and 6). The HRV appears to decrease as RMSX error increases during bank maneuvers.

Eyeblink (EOG)

Number of Blinks. The first test performed on the number of blinks was the flight segment by session block interaction ANOVA, which was not significant. The main effects of session block and day/night were also not significant. Flight segment was significant, $F(7,28) = 11.55$, $p < 0.0001$. As shown in Figure 14, subjects blinked more during take-off than during any of the other flight segments.

Another test was performed on the eyeblink data, similar to the heart rate. The first and last 30 seconds of data in each segment were obtained. The flight segment by time block (8 X 2) interaction was significant, $F(7,28) = 17.46$, $p < 0.0001$. As can be seen in Figure 15, subjects blinked more during the first 30 seconds of the take-off segment than during the last 30 seconds.

Blink Interval. None of the dependent variables affected blink interval, except for flight segment. The main effect of segment was significant, $F(7,28) = 2.68$, $p < 0.0297$. Blink interval was larger during the resting segment than during all other segments.

Half-Amplitude Closing Duration. There were no significant interactions or main effects for this eyeblink measure.

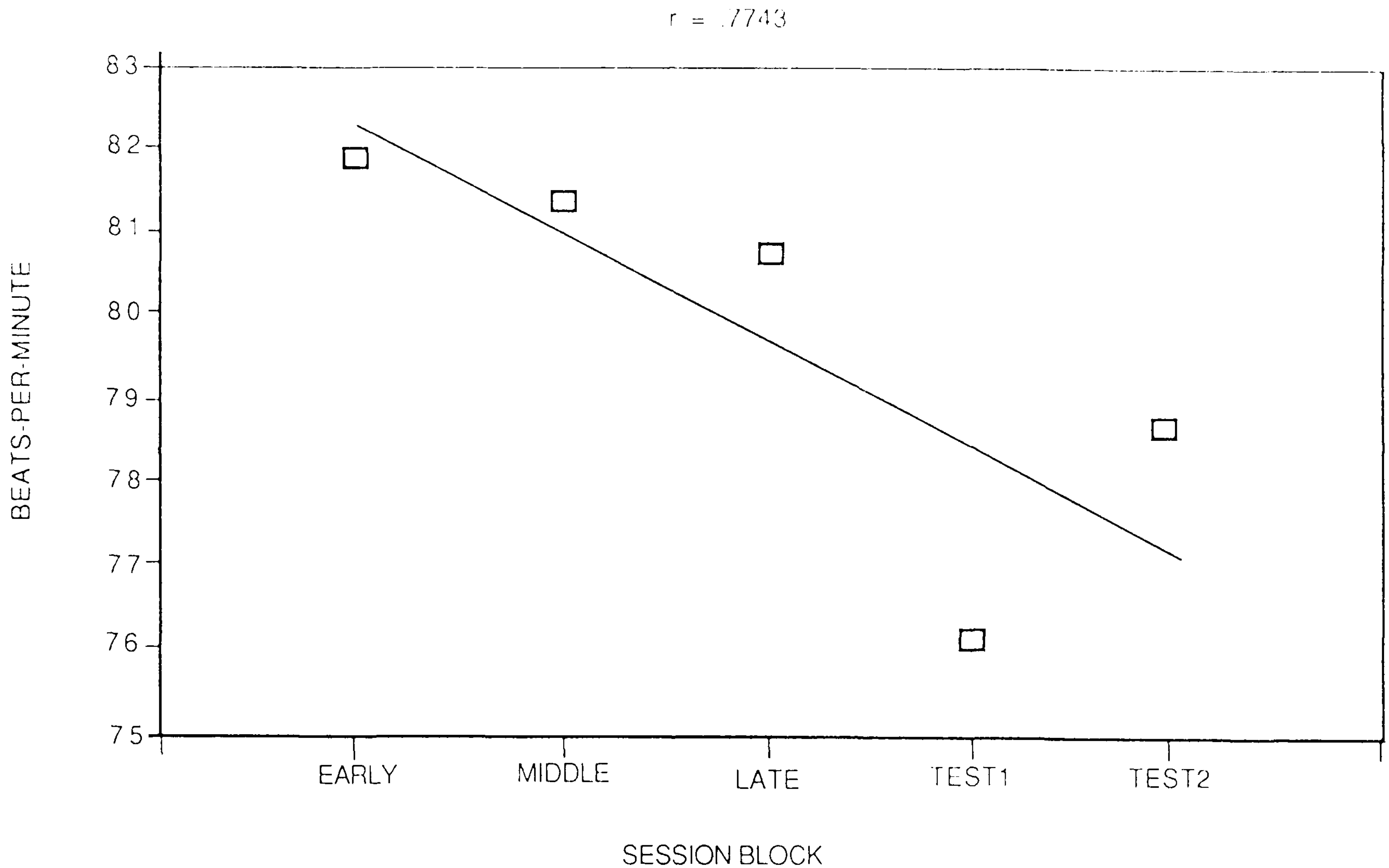


Figure 12. Beats Per Minute Regressed with Session Block

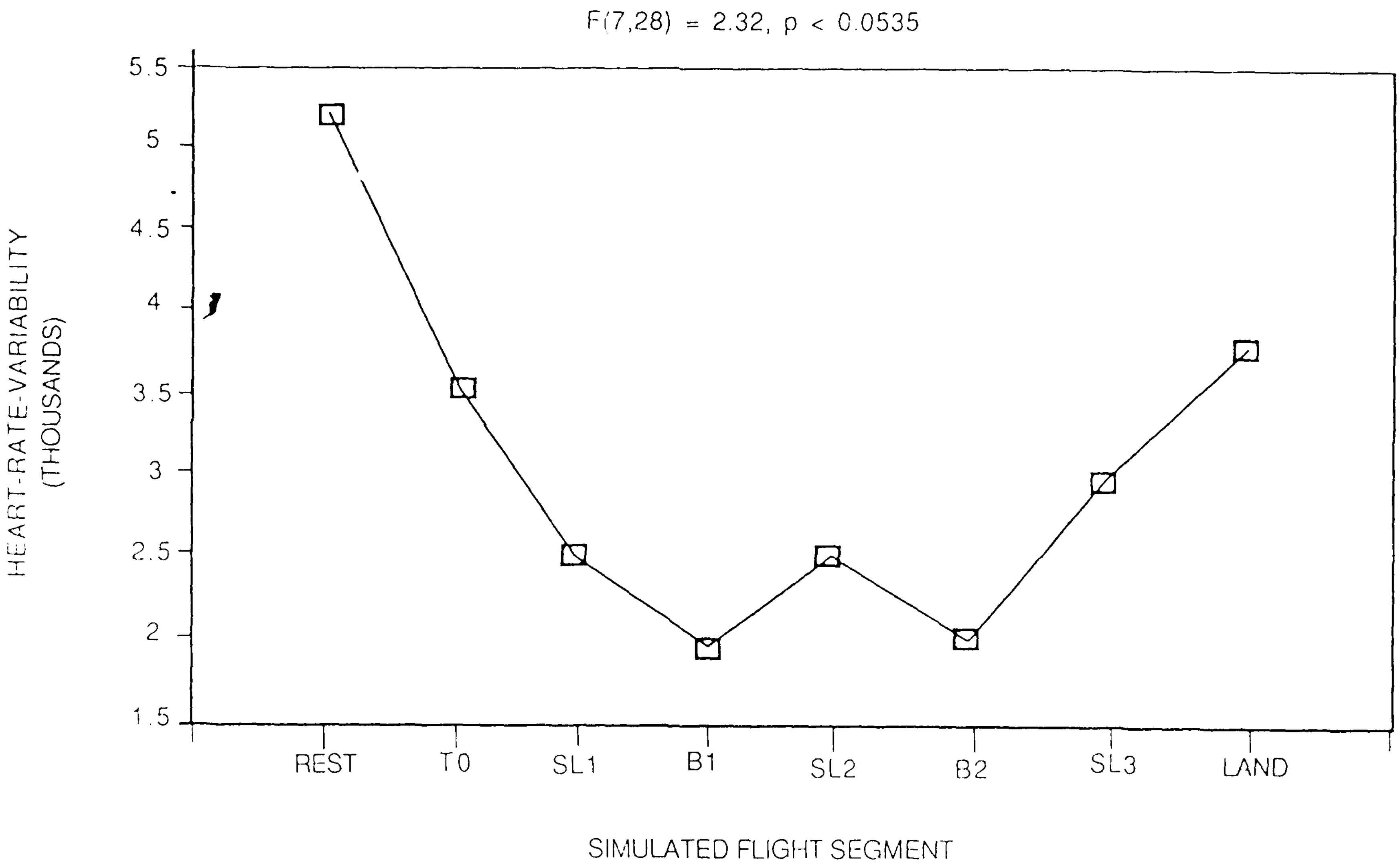


Figure 13. Heart Rate Variability Segment Effects

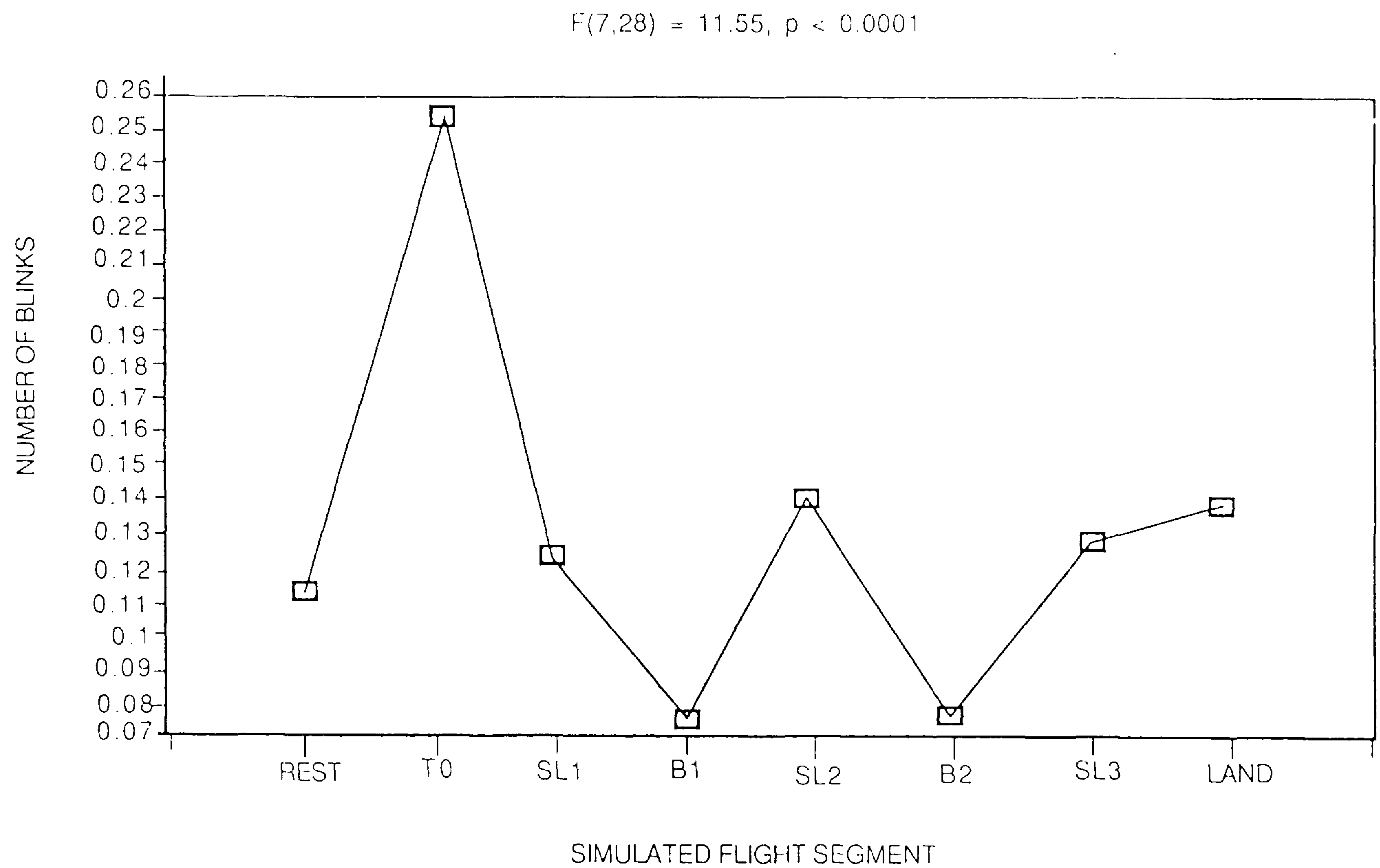


Figure 14. Blink Rate Segment Effects

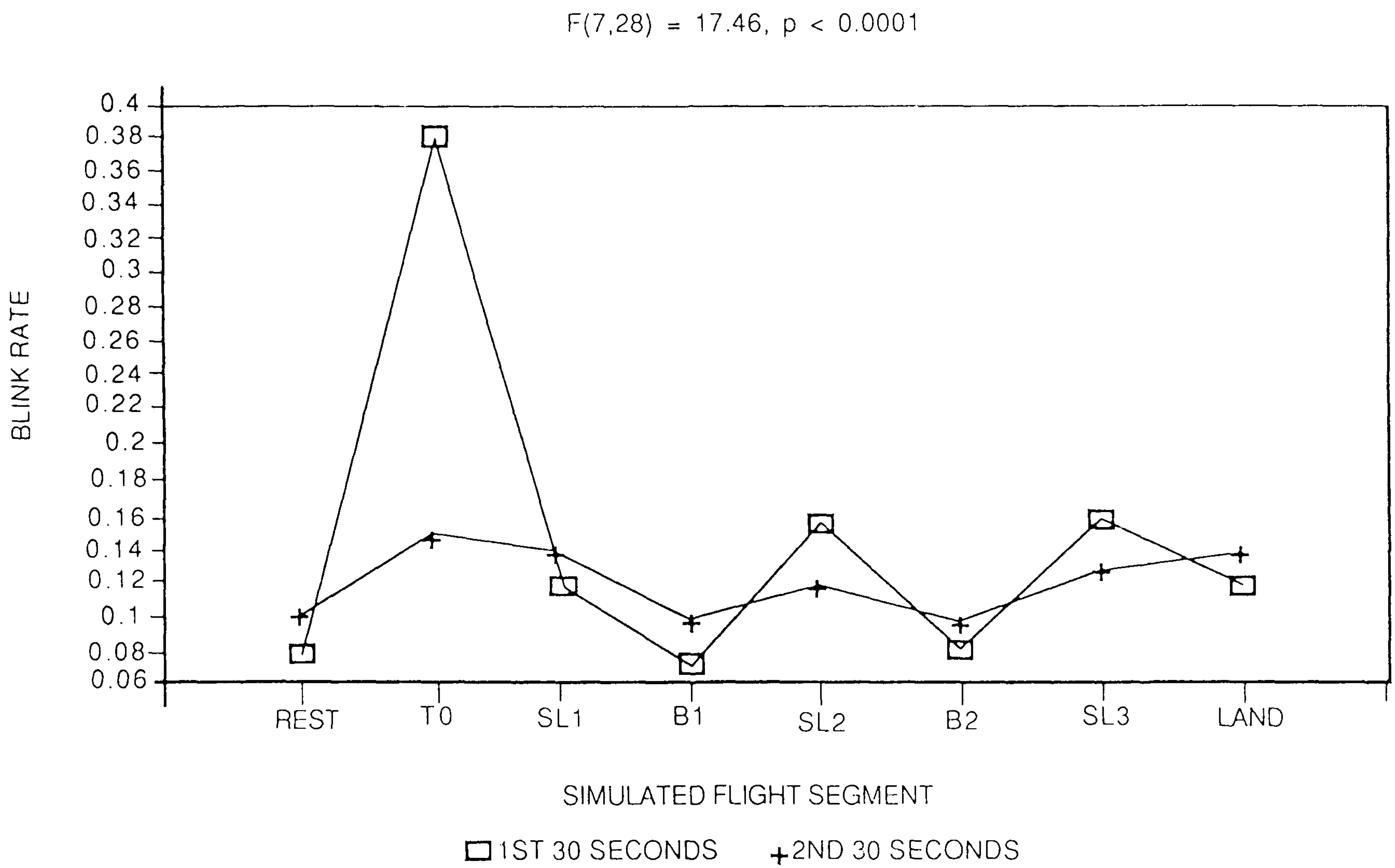


Figure 15. Blink Rate Segment Effects in First and Second 30-Second Blocks

Closing Duration. The interaction for flight segment by session block was not significant. However, all three main effects were significant: day/night, $F(1,4) = 11.18$, $p < 0.0287$; session block, $F(4,14) = 5.04$, $p < 0.0100$; and flight segment, $F(7,28) = 3.44$, $p < 0.0087$. As can be seen in Figure 16, closing duration was longer during day than during night flights. Also, closing duration was longer during both test conditions than during early, middle or late session blocks (see Figure 17). Closing duration was longer during the resting segment than during all other flight segments, as shown in Figure 18.

Evoked Potentials

The data reported below were taken from both the rare tone and frequent tone evoked potentials obtained from the oddball task. Components were selected according to latency and waveform criterion. For the identification of the P200 component, a positive-going deflection had to occur within a 150-300 msec time window after onset of the stimulus. For the N200 component, a negative-going deflection had to occur within 150-350 msec. For the P300, a positive-going component had to occur within 300-700 msec. In many instances there were missing components in the individual subjects' waveforms. For example, the frequent tone does not elicit a N200 or P300 for most subjects. The analysis for these components was not included since over half of the data points were missing. For the rare tone evoked potentials there were also missing data points. However, all components were included in the analysis since the amount of missing points was negligible. Representative subject rare tone EPs, as plotted by the NWTB, are presented in Figure 19. The P300 components on both averages are marked by the cursor. The latency and amplitude values of the P300 are given at the bottom of each waveform.

Rare Tone Evoked Potentials

The P200, N200 and P300 amplitude and latency values were all tested for significance in relation to day/night, flight segment and test block (see Table 3). The only component to show significance was the P200. P200 amplitude varied according to flight segment, $F(7,28) = 4.03$, $p < 0.0336$.

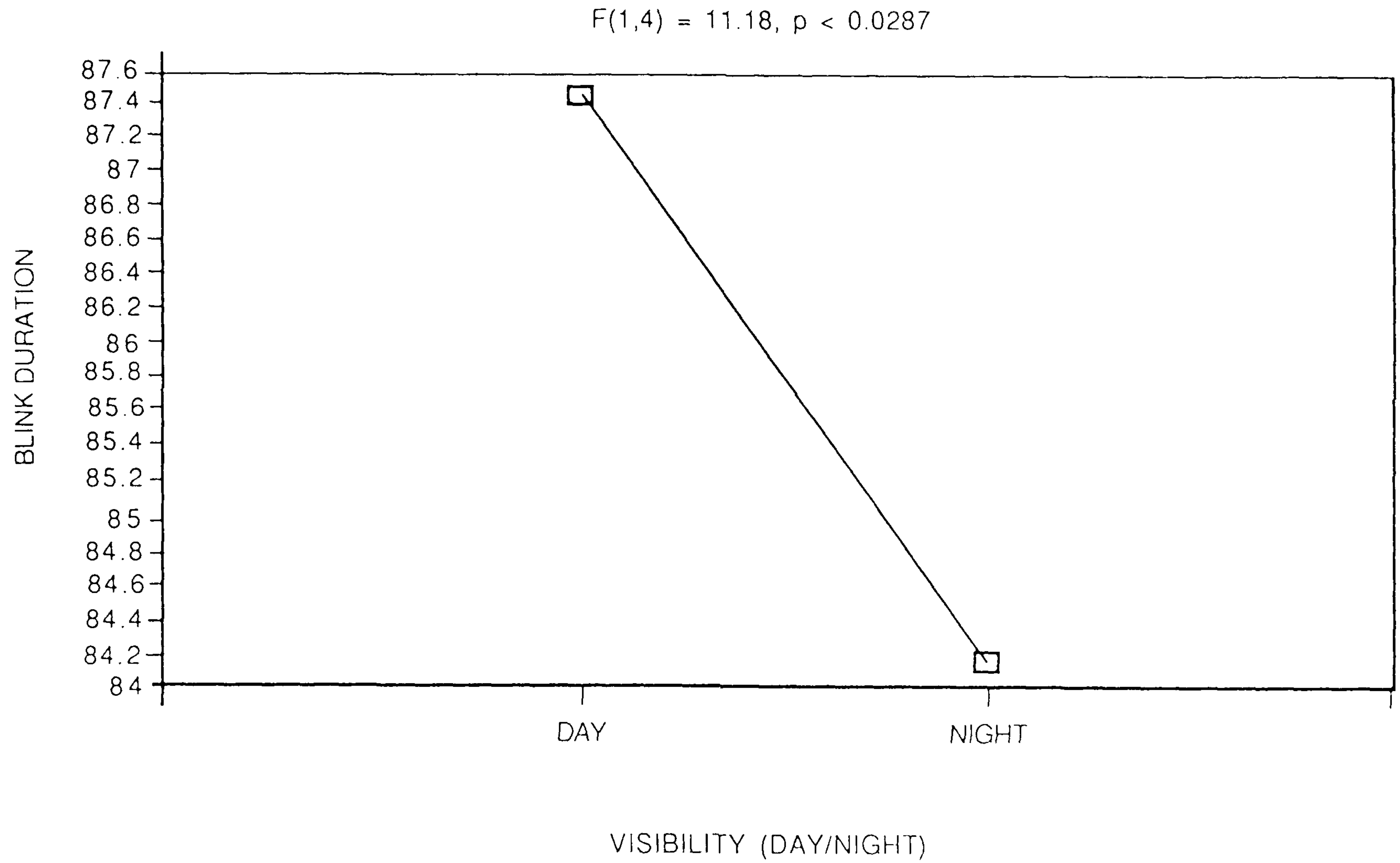


Figure 16. Blink Duration Visibility Effects

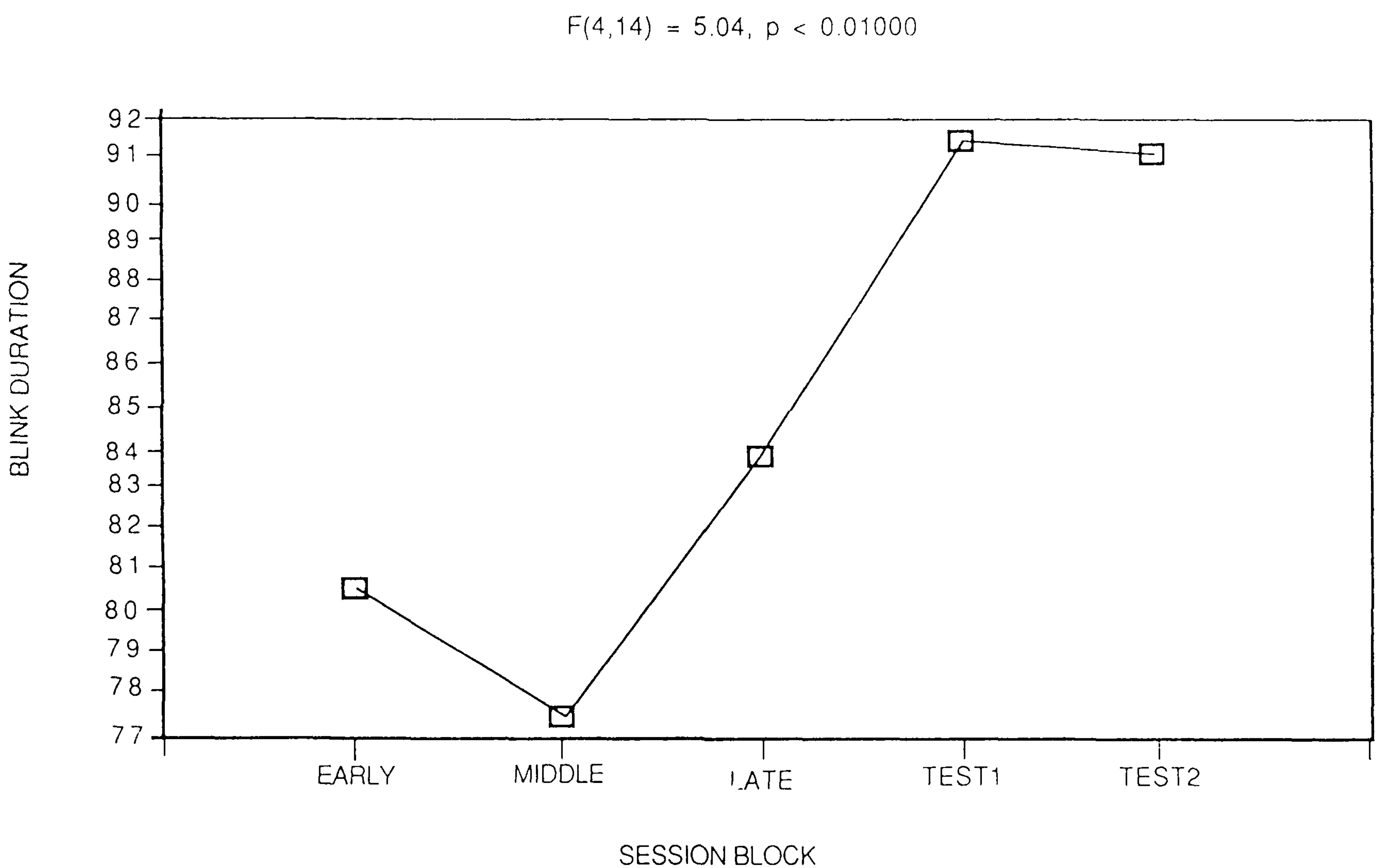


Figure 17. Blink Duration Session Block Effects

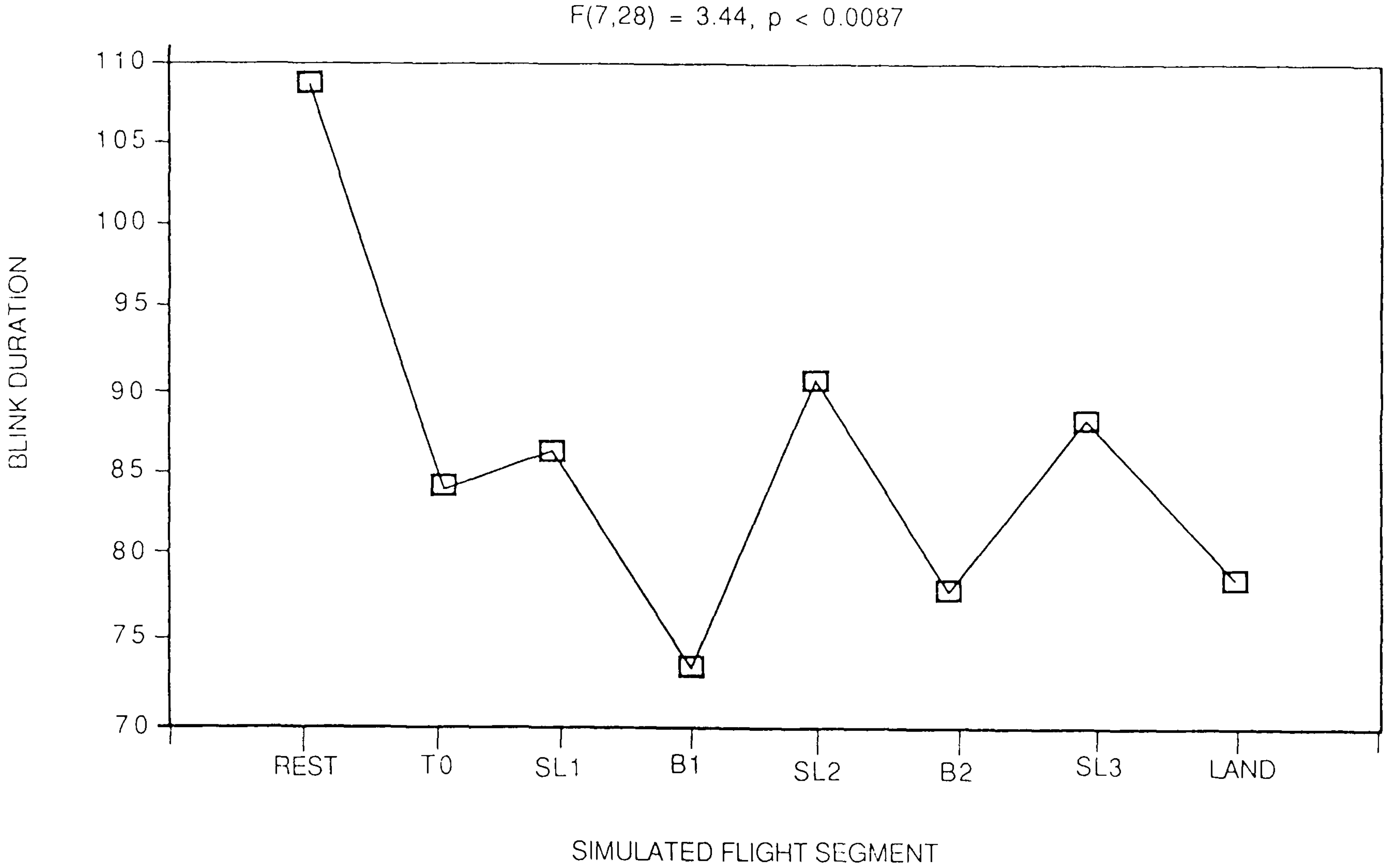


Figure 18. Blink Duration Segments Effects

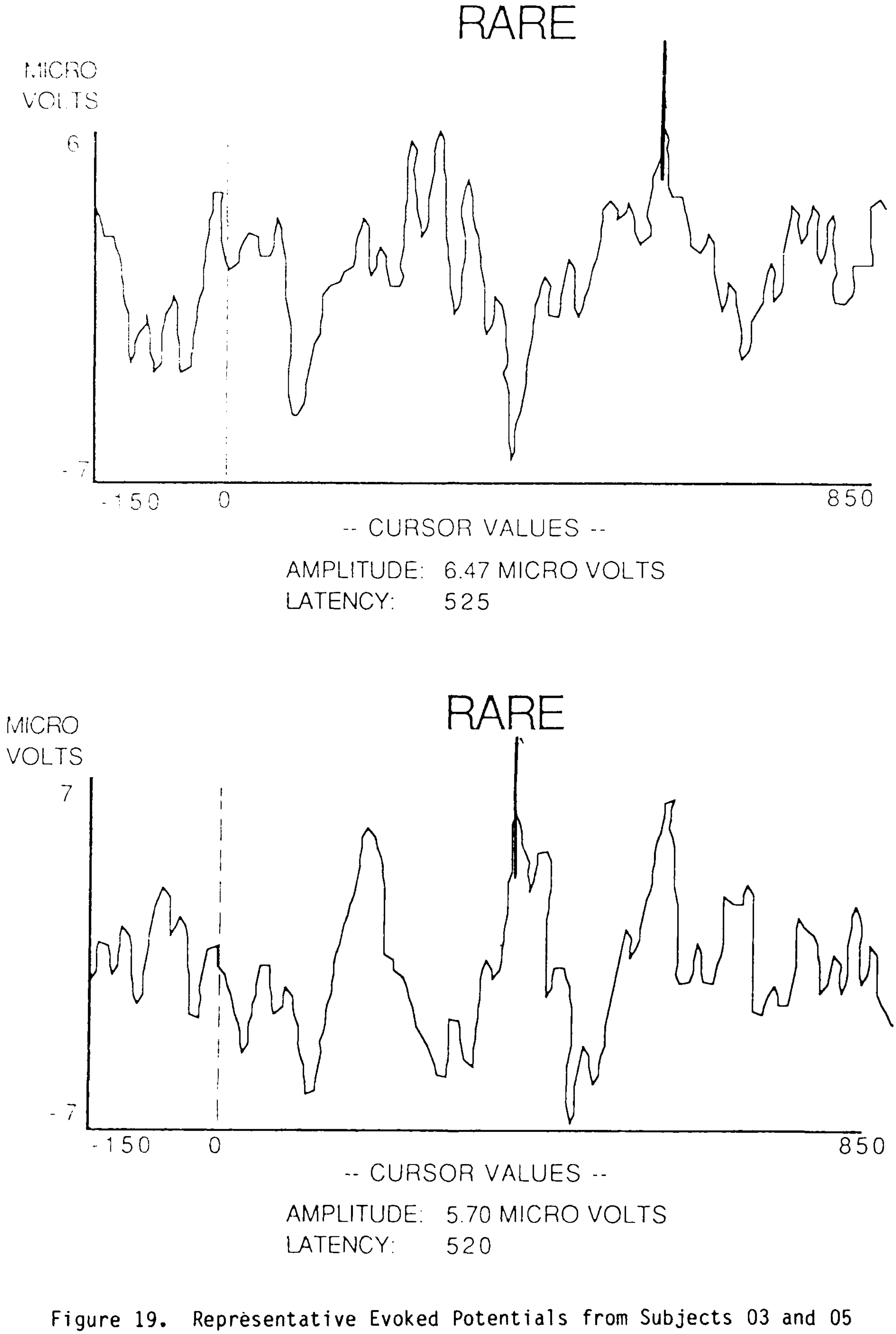


Figure 19. Representative Evoked Potentials from Subjects 03 and 05

TABLE 3. SIGNIFICANT TABLE FOR ALL EP COMPONENTS (P VALUE)

	P200 Amplitude	P200 Latency	N200 Amplitude	N200 Latency	P300 Amplitude	P300 Latency
RARE TONE COMPONENTS						
Day/Night	0.2058	0.9116	0.6989	0.7351	0.9502	0.4476
Segment	0.0036*	0.4529	0.6463	0.9698	0.1138	0.9630
Training Block	0.9858	0.1705	0.9222	0.4496	0.5112	0.3885
Day/Night X Segment	0.3804	0.1302	0.7005	0.1613	0.2955	0.3893
FREQUENT TONE COMPONENTS						
Day/Night	0.0550	0.5259				
Segment	0.0002*	0.3573				
Training Block	0.4101	0.9891				
Day/Night X Segment	0.7877	0.3254				

Amplitude = millivolts
Latency = milliseconds
* $p < 0.05$

Amplitude was larger during rest than during take-off, the first and second straight and level, and the first bank, which in turn were larger than the second bank, the third straight and level, and landing (see Figure 20).

Frequent Tone Evoked Potentials

The P200 amplitude of the frequent tone evoked potential also varied according to flight segment, $F(7,28) = 6.33$, $p < 0.0002$. As can be seen in Figure 21, amplitude was larger during rest and take-off than during all other segments of the flight. Day/night and test block effects were not significant.

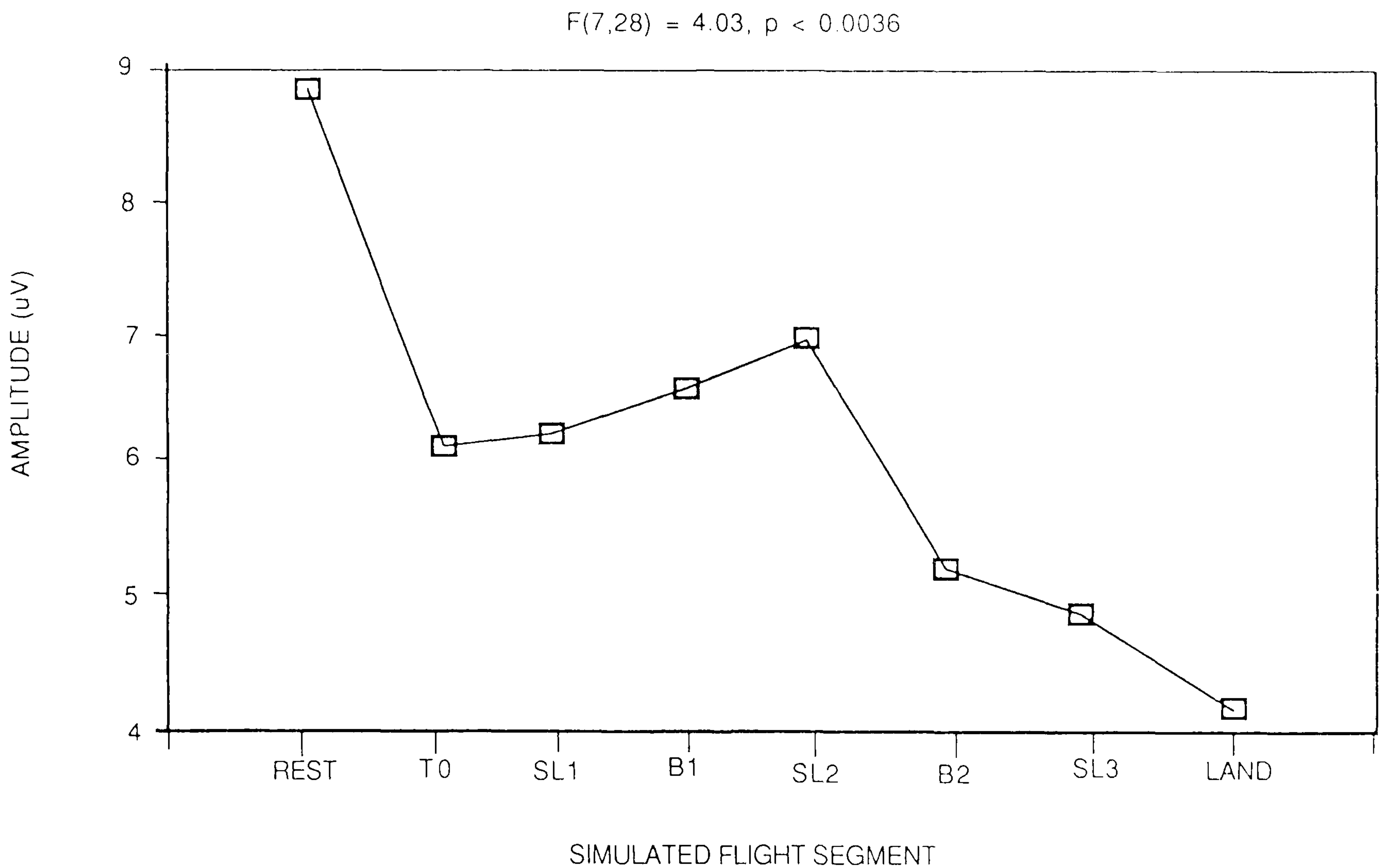


Figure 20. Rare Tone P200 Amplitude Segments Effects

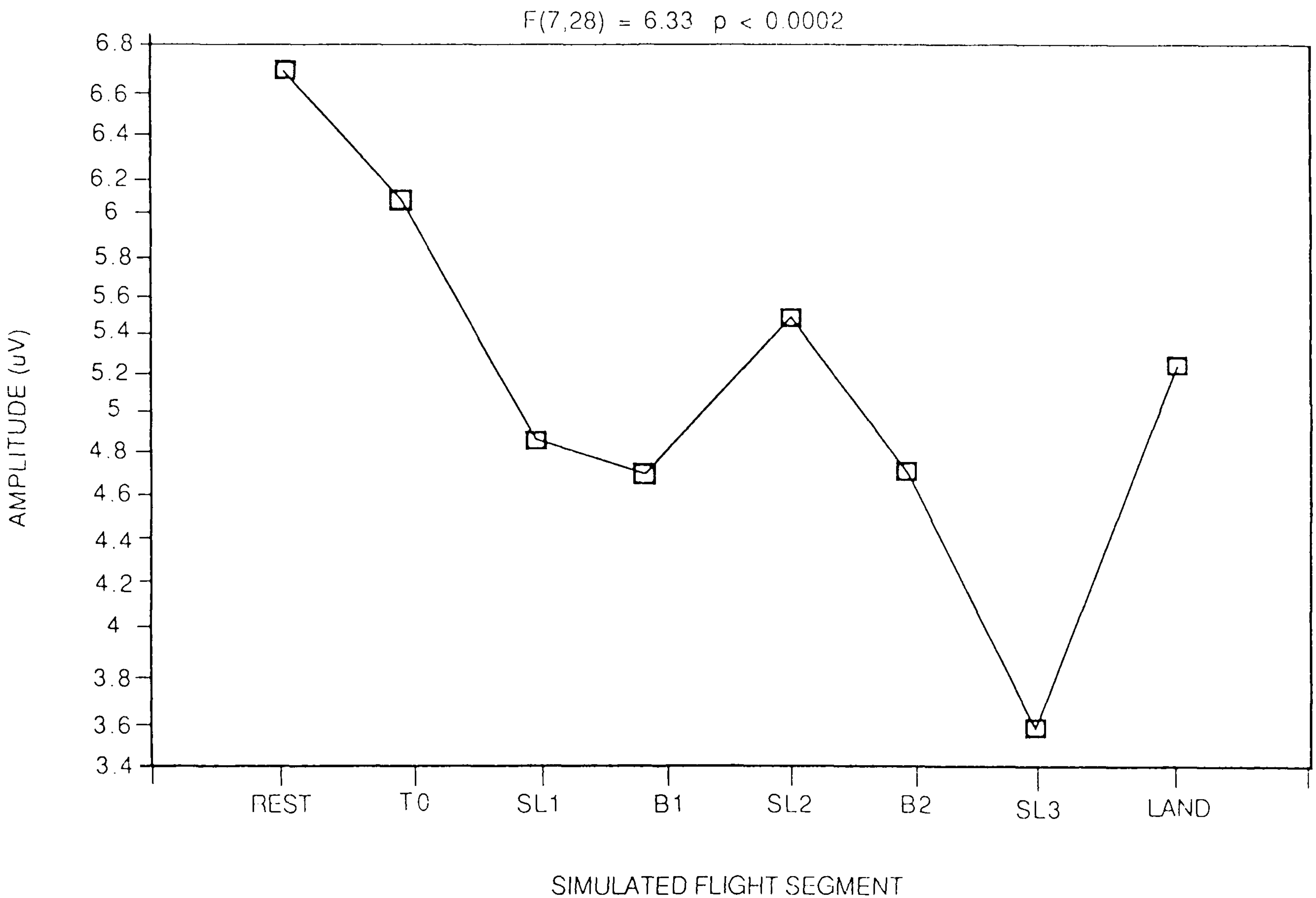


Figure 21. Frequent Tone P200 Amplitude Segment Effects

Section 6
DISCUSSION

The SG behavioral data showed clear session block effects in all three control axes. Subjects' error in controlling their aircraft and following the lead plane was larger at the beginning of the sessions. Also, heart rate decreased as familiarity and competence on this flight task increased. The convergence of the behavioral error and heart rate data suggests a definite training effect during these sessions.

There was an effect attributed to day/night visibility which was surprising in its direction. Error in the lateral offset and trail distance axes increased during day flight over that found with night flight. Subjects' comments as to the difficulty of the flights under these visibility conditions matched the increase in error during the day flight. Most subjects said that the extraneous display information during day flight interferred with the task of following the lead plane. Night flight removed this visual information, and added more "relevant" display cues such as lead tail/wing lights and engine burn as well as landing lights on the runway. Subjects also reported improved depth cues during night flight. This visibility factor also affected blink behavior in the manner suggested by Morris ((1985) see Appendix). The length of time the eye remained closed (blink duration) was longer when control error was greatest, i.e., during day flight.

It should be noted here that the flight simulation was a low fidelity simulation with unrealistic terrain and horizon cues. In light of this situation, it might not be too surprising to find converging behavioral and physiological data that suggests night flight was easier than day flight. These patterns would not be expected in a high fidelity simulation or real flight.

The effects of the different flight segments on control error were different for lateral offset and trail distance. The lateral offset error between the subject's plane and the lead was greatest during both the first and second bank maneuvers. Furthermore, both HRV and number of blinks

decreased during these two bank segments. Recall that increased load on an operator decreases HRV and number of blinks. The corresponding changes in these three measures (offset error, HRV, and blinks) strongly suggest that the first and second bank maneuvers were the most behaviorally difficult segments in the flight.

Trail distance error, however, was most affected only by the first bank maneuver. The data suggests that subjects were playing "catch-up" in trail distance to the lead from the first bank up until landing (see Figure 6).

An interesting aside is that, had only the error data in the three axis pursuit tracking task been obtained, it would be very difficult to say where in the flight the greatest control demand was imposed on the subjects. The altitude axis showed no segment effects, and the offset and trail axes show two distinctly different patterns of error. Without the physiological data, it would have been a toss-up between the offset and trail patterns in determining the segments with the greatest control load.

The evoked potential data were disappointing. The P300 components did not covary with any of the independent variables (segment, session or visibility). The rare and frequent tone P200 components did show differences between preflight baseline and all other segments of the flight. A decreasing rare tone P200 amplitude trend from the second straight and level to the landing segment was also apparent, albeit insignificant.

Other physiological results identified flight effects not found with the evoked potential or behavioral data. Heart rate in BPS showed increases during the landing segment, specifically during the last 30 seconds where subjects were actually touching down on the runway. At least during the last few sessions, if the subjects were going to lose control of their aircraft and possibly crash, it would have been during this segment. The heart rate data reflect the heightened arousal during landing that the other measures do not.

The eyeblink data showed that number of blinks were greatest during take-off, specifically during the first 30 seconds of the segment where subjects

were heavily scanning the cockpit instrumentation displays. This portion of the segment was spent readying the airplane for take-off (e.g. pretake-off checks, such as wing flap position) prior to the lead fly-by. The last 30 seconds of the segment were actual take-off after the lead fly-by and during this portion the number of blinks is the same as the rest of the segments.

In summary, the physiological data provided information about the simulated flight task that not only corresponded to and further clarified the behavioral data, but also showed differences between segments of the flight task that would not have been apparent with the behavioral data alone.

Section 7
RECOMMENDATIONS

The three objectives outlined in Section 2 of this document included a test of the interface between the simulator and the NWTB, determining the value of the physiological data collected and formulating recommendations for further use of these measures in a larger simulation effort.

As shown elsewhere in this report, the problems of interfacing the simulator and NWTB were time consuming, complicated and difficult to solve. Yet there is now in place the capability to record physiological data in conjunction with SABER simulator events. The interface problems have been solved during this preliminary effort. However, as discussed in Section 3, the NWTB still possesses user-intensive requirements for obtaining data means and summary statistics. Without a substantial investment in software changes these requirements will remain. It now becomes a question of a trade-off between the cost of employing an NWTB operator, the intrinsic worth of the data obtained, and delays in obtaining reduced/analyzed data.

From the data reported above, physiological results not only clarified behavioral issues, but also provided additional information not obtained otherwise. This simulation was a low fidelity, "video game" flight task, and yet the heart rate and eyeblink measures were sensitive to flight changes. With the exception of the evoked potential oddball task, these physiological measures added a wealth of information to the evaluative data base. It would be expected that with a more realistic simulation these measures would add the same, if not more, dimensionality and precision to any flight evaluation.

As to the failure of the evoked potential technique, there are two possible methodological changes that could be implemented in the future to increase measurement sensitivity. The first would be to attain an eyeblink correction program to ensure a larger number of single trials in each of the subjects' averages. The other would be to change the evoking stimulus from a

secondary oddball paradigm to a relevant flight task. Any flight event that is repeated throughout the course of the flight mission could be used to time-lock and average the EEG.

An important question for the future use of the SABER simulators would be that of practice and transfer of learning. The physiological measure most amenable to this has been heart rate (as reported in the literature and as evidenced above). The subjects used in this study were not experienced pilots. Pilots with experience would be expected to give substantially different response profiles, not only behaviorally but physiologically. It is highly recommended that any SABER investigation of experts versus novices, as well as responses obtained from simulation versus actual flight, include physiological measures.

Other recommendations for a larger, high-fidelity simulation would be:

1. A priori identification of the flight task segments of interest, and timing of these segments to the nearest second (requires the simulation be in place with existing time-line analyses).

2. Allow for the marking of the simulator's and NWTB's time history for such things as pre-programmed emergencies and unplanned crashes, etc.

3. Obtain an operator trained on the NWTB and simulator parameters before data collection, as well as a consultant for the design of physiological methodology.

Overall, it is recommended that physiological response measures be obtained during any simulator mission. In part-task through whole mission scenarios these measures can only add to the understanding of the crewmember's role during flight.

Section 8
CONCLUSIONS

The SABER facility has been established to investigate the issues involved in developing advanced crew systems. The development of these new systems will undoubtedly open up new dimensions of information transactions at the point of crewmember control and systems displays. The ultimate goals of the SABER facility include maximizing the effectiveness of the crewmember within the system, while minimizing the number of crewmembers needed for the successful performance of a mission.

The SABER facility is particularly suited for the investigation of evolving and new strategic missions. The technologies needed for the successful accomplishment of these missions will be a significant driving factor concerning crewmember interface design, especially in such areas as the offensive weapons crewstation. The introduction of technologies new to the crewstations, such as complex artificial intelligence modules, will provide quantities of unknown and indeterminate information to the crewmember. It will become increasingly important, and increasingly difficult to investigate crewmember workload and situational awareness.

The SABER facility is employing a three-pronged approach in documenting both crew workload and situational awareness. The first is the collection of relevant time based behavioral data during real-time simulated missions, utilizing the four-member SABER crewstation. Data such as flight control parameters, weapons delivery timing and accuracy, and the timely response to task relevant information are currently available to provide an objective data base.

The second is the collection of each crewmember's ratings of the difficulty of the mission segments. Data can be collected during relevant points of the mission simulation by using such techniques as the Subjective Workload Assessment Technique (SWAT) via the simulator's intercom system. Subsequent debriefing of the crewmembers using recall probes augmented with behavioral data can add to the subjective data base.

The third technique will be a class of physiological measures utilizing the NWTB. The NWTB is a computer based collection tool that has undergone specific modifications for SABER requirements. The modifications include expansions to memory that allow data collection from two crewmembers simultaneously for up to 6 hours in duration. Interfaces that allow for the mutual time synchronization of the NWTB and SABER were developed which, in turn, allow for the correlation of physiological, behavioral, and subjective data bases. The collection of physiological data during a simulated flight mission was shown to be a successful and useful technique for the investigation of crewmember workload.

Utilizing this three-pronged approach of behavioral, subjective, and physiological measures should provide a multidimensional picture of workload and situational awareness. Now that the SABER facility possesses the capability of collecting all three measures, it is expected that the investigations of crewmember/crewstation interactions will be a truer representation of the actual mission requirements involved.

APPENDIX

In 1987, an AGARDograph (edited by Stan Roscoe) was devoted entirely to outlining the state-of-the-art techniques used to identify and quantify the characteristics of any given flight system (AGARD No. 282, 1987). In 11 of the 16 chapters dealing with workload and human-machine interactions, psychophysiological techniques were utilized.

The multidimensional qualities of workload have recently been recognized, as evident in the continuing interest of using a three-pronged approach to investigate human-machine interactions. This approach consists of using objective, subjective and physiological techniques.

Most investigations of new or existing crew stations, cockpits and assembly lines use the objective category of techniques. This category includes detailed time-line analyses of task demands, time and motion studies, and performance measures (including secondary task techniques). Furthermore, the use of subjective techniques, such as rating scales obtained from operators during performance, have proven reliable as indicators of the level of perceived workload.

Psychophysiological techniques are relatively new in the applications field and so do not benefit from repeated use as do the more routinely used methods of objective and subjective techniques. However, there exists a wide body of laboratory data that supports the idea that much can be gained from these physiological techniques, as well as the few reports that deal with data obtained from simulations and actual in-flight operational missions. Psychophysiological techniques can fill in the informational gaps left by objective and subjective techniques. Time-line analyses yield information only about the task, and performance measures give information only about human-machine status at that given time. It does not answer questions about predictive performance under stress, such as emergencies and bottlenecks in human capacity (O'Donnell and Eggemeier, 1986). Subjective techniques may answer some of these questions, yet it has been pointed out before that these techniques have their own unique shortcomings, including individual biases, poor individual replication under

similar circumstances, and delayed ratings under high workload conditions (Hart, 1982). Psychophysiological techniques have been used precisely to fill in these informational gaps. It is recommended that this three-pronged approach (objective, subjective, and physiological) be used to obtain a more complete picture of human-machine interactions.

PSYCHOPHYSIOLOGY IN RELATION TO FLIGHT PARAMETERS

Heart Rate (or ECG)

As early as 1967, researchers were examining changes in the cardio-vascular system of pilots during different flight segments. Smith (1967) reported elevations of heart rate during take-off and landing phases of commercial airline routes. Roman, Older and Jones (1967) reported that navy pilots flying combat missions had higher heart rates during carrier landing than during actual target acquisition and weapons delivery. This finding opposed the belief that acquisition and delivery was the most taxing and dangerous flight segment.

Increased heart rate during take-off and landing segments of any given commercial flight is not surprising since flight time between these two segments does not usually entail a large amount of pilot activity. However, in the above military scenario, the segments between take-off and landing required at least the same amount, if not greater, pilot activity and attention to flight parameters as during take-off and landing. This psychophysiological technique revealed information not obtainable with the more traditional measures.

The usefulness of psychophysiological measures was also demonstrated in a recent application of heart rate measures during the performance of an A-7D mission scenario. Skelly, Purvis and Wilson (1987) and Wilson, Purvis, Skelly, Fullenkamp, and Davis (1987) reported data taken from pilots during the training performance of a tactical mission in both actual and simulated flight. Heart rate was sensitive to changes in mission segments as well as in-flight versus simulator differences. Specifically, mission events affected heart rate in the lead and wing but not in the simulator. For the

lead flights. pilots' heart rate was higher during pretake-off, take-off, guns jink, 90-degree slice maneuver, weapons delivery, and landing than during briefing, fly-over update, and cruise. For the wing, pilots' heart rate was higher during these same segments than for briefing and cruise. Pilots' heart rates during simulator flight did not show these differences. Furthermore, heart rate was higher overall during the lead flight than during the wing flight, and higher during wing than during simulator flight. Wilson et al. (1987) suggested that the added workload of piloting an actual aircraft over that of a simulator, and the increased duties/ responsibilities of the lead pilot, led to these increases in heart rate.

These results lend support to the assumption that simulations of flight missions will not elicit the same physiological responses from pilots that are found during actual flight. This has direct bearing on heart rate taken during simulator flight. As discussed in Skelly et al. (1987), there is a need to measure physiological reactivity in simulators to examine how and where simulator fidelity affects crew performance. It may be that the mere physical similarity between simulators and aircraft is not sufficient to have true applicability in training, licensing and certification, and general engineering research. In the SABER laboratories, however, heart rate responses obtained from simulation can be used as baselines in relation to actual aircraft flight. Furthermore, heart rate can be applied as a measure of "relational" fidelity. As an example, two separate weapons delivery simulations may elicit different increases in heart rate.

Related results obtained in the laboratory suggest further uses of heart rate in examining training effects. In 1973, Zwaga asserted that heart rate reactivity was directly related to the amount of time any given subject had practiced on a task. The main hypothesis was that the initial introduction to a task precipitated increased heart rate, and that further performance of that task would see gradually decreasing heart rate. These effects occur even when behavioral performance is stable from the beginning to the end of the task (McCloskey, 1987). This has importance when training and task load effects occur in a situation at the same time. In other words, task load effects may be superimposed over a gradually

decreasing trend in heart rate, and this trend may even obscure the task load effects.

Eyeblink (or EOG)

Eyeblink behavior during demanding tasks has been characterized as follows:

- As attentional demands increase, the number of blinks decreases.
- As attentional demands increase, the length of time the eye is closed decreases.
- As fatigue increases, number of blinks and duration of blinks increases.

These phenomena have been heavily documented in the literature (Bauer, Strock, Goldstein, Stern, and Walrath, 1984; Stern and Skelly, 1984; Stern, Walrath, and Goldstein, 1984). Stern, Walrath, and Goldstein (1984) reviewed the evidence concerning eyeblink behavior during task performance. It is postulated that blinking is best understood in terms of a cognitively-based mechanism which suppresses blink behavior until such times as decisions about external stimuli have been made.

Stern and Skelly (1984) reported data that were obtained during high-fidelity simulator missions. Blink rate discriminated between the pilot in control of the aircraft and the person acting as copilot. Furthermore, mission segments affected blink rate. During the weapons delivery and "coping with threat" segments, blink rate decreased. These mission segment effects on blink rate are evidence that blink rate can be a measure of task demands in an aircraft setting. Also, blink duration increased from pilot to copilot, and was shown to increase in latency as time-on-task increased.

Another interesting aspect of blink behavior is evidence that high rates of long-duration blinks during the "taking-in" and processing of relevant stimuli may be an indication of erroneous performance (missed signals, incorrect responses, etc.). Morris (1985) used a basic instrument flight

trainer with cockpit motion and engine sound simulation to examine eyeblink effects in relation to piloting errors. Eyeblink results showed that the longer the closure duration, the higher the error in airspeed, heading and altitude during the prespecified flight course. Furthermore, as the frequency of long duration blinks increased (those longer than 500 msec in duration), error also increased. Morris (1985) further suggested that these eyeblink measures can be used to predict mistakes in flight performance. Errors in airspeed, heading and altitude were preceded by increases in blink duration far enough in advance to become candidates for the valid prediction of "catastrophic failure." Eyeblink changes occurred before performance decrements were found.

The above results document the use of eyeblink measures in simulation and actual flight. These measures proved useful in obtaining information about pilot versus copilot attentional demands, task loads, and investigations of EOG as predictors of flying performance decrements.

Electroencephalography (or EEG)

Perhaps the most intuitively appealing physiological measure is that of brain activity. If the cognitive demands on an operator are of interest, surely the on-going activity of the brain must offer insight into mental workload. In practice, however, it has been found that the EEG does not offer this insight easily. Difficulties arise with detecting the extremely small EEG signal (on the order of 5-20 microvolts) from the human scalp. In some cases amplification of the EEG must be as large as 100,000 times the original signal. When this problem is overcome, EEG techniques can be useful (O'Donnell, 1979).

On-going EEG obtained during actual aircraft and simulator flights has been submitted to spectral analysis and correlated with mission event as well as flight formation position (Skelly et al., 1987; Wilson et al., 1987). The 4-7.5 Hz band, the 8-13 Hz band, the 14-19.5 Hz band and the 20-30 Hz band of the EEG spectra showed increased activity during the lead and wing aircraft flights over that found during the simulator flight. Furthermore,

during high-G stress in the aircraft flights, the 8-13 Hz band of the EEG showed heightened activation.

While these results are promising, Wilson, et al. (1987) cautioned against too strong of an interpretation. Problems with obtaining the EEG signal stemmed from the electrically noisy environment of the aircraft (coupled with movement artifact) not found in simulation. This increased noise could easily have caused the increased activity of the EEG signal, unrelated to any cognitive functions of the pilots. Wilson, et al. (1987) recommended changes in EEG collection techniques, specifically larger amplification of the EEG signal closer to the point of the electrical source. Presently, stronger preamplifiers located within inches of the scalp contacts are being tested on pilots flying F-4 aircraft.

When these problems of signal noise have been controlled, spectral analyses of on-going EEG signals have proven to be generalized measures of arousal. Sterman, Schummer, Dushenko, and Smith (1987) found that 8-11 Hz amplitude discriminated between pilots during resting and when flying an airplane. Furthermore, sensorimotor cortex scalp sites showed higher amplitude responses in the 8-11 Hz band during controlled flight tasks than the visual cortex scalp sites. These results point out the usefulness of on-going EEG measures, especially when used to investigate differences between aircraft and simulation flights, and task/no-task situations.

On-going EEG does not lend itself well to answering more specific questions about flight tasks. A technique known as the averaged evoked potential (EP) has been used extensively in the laboratory to answer very specific questions about task loads. The technique consists of taking one second "snap-shots" of the on-going EEG potentials that occur in response to discrete stimuli, and then superimposing these snapshots one upon the other to obtain an average EP. The technique is limited by the requirement of multiple occurrences of the discrete "evoking" stimuli to obtain multiple snap-shots for the average EP. The recommended number of snap-shots required to obtain a reliable average EP has ranged from 10 to 100 (Wickens, Kramer, Vanasse, and Donchin, 1983). In some situations, the evoking stimuli do not occur often enough to obtain a "reliable" average.

For example, Albery (1988) obtained EPs elicited from both a primary and secondary task while pilots were under centripetal force in a high-G simulator. The number of snap-shots available for the EP averages ranged from 1 to 22 (mean=11, std=3). Albery (1988) identified two major problems that caused many of these single trial snap-shots to be lost: (1) excessive eyeblink contamination during high-G stress, and (2) electrical noise generated by the motion-based centrifuge apparatus. It is interesting to note, however, that results obtained from these EP averages still matched the a priori hypotheses of the author. The components of the EPs to the primary tracking task increased in amplitude and latency as the difficulty of the task increased. Furthermore, the components of the EPs to the secondary target acquisition task increased in latency as the number of targets increased. This is evidence that, even though EP signals may be noisy and contaminated by artifacts in motion-based simulators and actual flight, this measure can still be useful.

In a more ideal situation, Kramer, Sirevaag, and Braun (1987) obtained EPs elicited by a secondary "oddball" task during the performance of a simulated flight mission. The simulation apparatus was fixed-based and did not cause large EP signal contamination, and any eyeblink artifacts could be mathematically corrected by a technique reported by Gratton, Coles and Donchin (1983). When the difficulty of the primary flight task was increased by introducing turbulence, subsystem failures, and increasing wind speed, the components of the EPs elicited by the secondary task (oddball tones) decreased in amplitude.

Since the SABER laboratories contain a fixed-based simulation capability, the electrical noise inherent in actual airplane cockpits and motion-based simulators should not be a problem. Evoked potentials should be relatively artifact free and easy to obtain. The secondary task discussed above (the oddball) seems to be a good candidate for indexing the difficulty of flight tasks. Also, eliciting EPs with discrete occurrences of task relevant stimuli, such as low-level emergency tones, could prove useful in evaluating flight tasks. Finally, on-going EEG spectral analyses have been used to determine overall arousal levels of the operator.

REFERENCES

AGARD (1987). The Practical Assessment of Pilot Workload. Stan Roscoe (Ed.), AGARD-AG-282, NATO Flight Mechanics Panel.

Albery, W. B. (1988). The Effects of Biodynamic Stress on Workload in human Operators. echnical Memorandum AAMRL-TR-88-004, AMD AFSC, Wright-Patterson Air Force Base, Ohio.

Bauer, L. O., Strock, B. D., Goldstein, R., Stern, J. A., and Walrath, L. C. (1984). Auditory discrimination and the eyeblink. Psychophysiology, 22, pp. 636-641.

Gratton, G., Coles, M. G. H., and Donchin, E. (1983). A new method for off-line removal of ocular artifact. Electroencephalography and Clinical Neurophysiology, 55, pp. 464-484.

Hart, S. G. (1982). Theoretical basis for workload assessment research at NASA Ames Research Center. In Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, AFFTC-TR-82-5.

Jasper, H. (1958). The ten-twenty electrode system of the international federation. Electroencephalography and Clinical Neurophysiology, 10, pp. 371-375.

Kramer, A. F., Sirevaag, E. J., and Braun, R. (1987). A psychophysiological assessment of operator workload during simulated flight missions. Human Factors, 29, pp. 145-160.

McCloskey, K. (1987). Evaluating a spatial task: Behavioral, subjective, and physiological correlates. In Proceedings of the 31st Annual Meeting of the Human Factors Society, pp. 774-778.

Morris, T. L. (1985). Electrooculographic indices of changes in simulated flying performance. Behavior Research Methods, Instruments and Computers, 17, pp. 176-182.

O'Donnell, R. D. (1979). Contributions of Psychophysiological Techniques to Aircraft Design and Other Operational Problems. AGARD-AG-244, NATO Aerospace Medical Panel.

O'Donnell, R. D. and Eggemeier, F. T. (1986). Workload assessment methodology. In Handbook of Perception and Human Performance, (Boff, K. R., Kaufman, L., and Thomas, J. P., Eds.), John Wiley and Sons: New York, pp. 42-1 to 42-49.

Roman, J., Older, H., and Jones, W. L. (1967). Flight research program: VII. Medical monitoring navy carrier pilots in combat. Aerospace Medicine, 38, pp. 133-139.

SAS User's Guide: Statistics (1985). SAS Institute Inc., (5th Edition), Cary, North Carolina.

Skelly, J. J., Purvis, B., and Wilson, G. F. (1987). Fighter pilot performance during airborne and simulator missions: Physiological comparisons. In Electric and Magnetic Activity of the Central Nervous System: Research and Clinical Applications in Aerospace Medicine, AGARD No. 432, pp. 23-1 to 23-15.

Smith, H. P. R. (1967). Heart rate of pilots flying aircraft on scheduled airline routes. Aerospace Medicine, 38, pp. 1117-1119.

Sterman, M. B., Schummer, G. J., Dushenko, T. W., and Smith, J. C. (1988). Electroencephalographic Correlates of Pilot Performance: Simulation and In-Flight Studies. Unpublished manuscript. Sepulveda Veterans Administration Medical Center, Sepulveda, California.

Stern, J. A. and Skelly, J. J. (1984). The eyeblink and workload considerations. In Proceedings of the 28th Annual Meeting of the Human Factors Society, pp. 942-944.

Stern, J. A., Walrath, L. C., and Goldstein, R. (1984). The endogenous eye blink. Psychophysiology, 21, pp. 22-33.

Wickens, C. D., Kramer, A., Vanasse, L. and Donchin, E. (1983). Performance of concurrent tasks: A psychophysical analysis of the reciprocity of information-processing resources. Science, 221, pp. 1080-1082.

Wilson, D. L. and Kuperman, G. G. (1988). Strategic avionics battle-management evaluation and research. Proceedings of the 40th IEEE National Aerospace Conference, Dayton, Ohio.

Wilson, G. F. and O'Donnell, R. D. (1988). Measurement of operator workload with the Neuropsychological Workload Test Battery. In Human Mental Workload (P. Hancock and N. Meshkati, Eds.), Amsterdam, The Netherlands: North Holland.

Wilson, G. F., Purvis, B., Skelly, J. J., Fullenkamp, P., and Davis, I. (1987). Physiological data used to measure pilot workload in actual flight and simulator conditions. In Proceedings of the 31st Annual Meeting of the Human Factors Society, pp. 779-783.

Zwaga, H. J. G. (1973). Psychophysiological reactions to mental tasks: Effort or stress? Ergonomics, 16, pp. 61-67.

U.S. GOVERNMENT PRINTING OFFICE 1988 - 648-002/80229